AD_____

GRANT NUMBER DAMD17-94-J-4149

TITLE: A Novel 80kDa Matrix-Degrading Proteinase in Breast Cancer Invasion and Metasis

PRINCIPAL INVESTIGATOR: Yuenian E. Shi, Ph.D.

CONTRACTING ORGANIZATION: Long Island Jewish Medical Center
New Hyde Park, New York 11042

REPORT DATE: September 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Annual (1 Sep 95 - 30 Aug 96)

4. TITLE AND SUBTITLE
A Novel 80kDa Matrix-Degrading Proteinase in Breast Cancer Invasion and Metasis

5. FUNDING NUMBERS
DAMD17-94-J-4149

6. AUTHOR(S)
Yuenian E. Shi, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Long Island Jewish Medical Center
New Hyde Park, New York 11042

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

Project 1. Recently, we cloned and characterized a novel human tissue inhibitor of metalloproteinases-4, TIMP-4. To determine if TIMP-4 can modulate the *in vivo* growth of human breast cancers, we transfected a full-length TIMP-4 cDNA into MDA-MB-435 human breast cancer cells and studied the orthotopic growth of TIMP-4-transfected vs control clones in the mammary fat of athymic nude mice. Overexpression of TIMP-4 significantly inhibited tumor growth rates; reached (3.5-14)-fold smaller primary tumor volumes and at sacrifice ($p < 0.01$); and gave lower rates of axillary lymph node and lung metastasis as compared with control clones. **Project 2**. A high-throughput direct differential cDNA sequencing approach was employed to identify genes differentially expressed in normal breast as compared with breast cancer. Of many putative differentially expressed genes, a breast cancer specific gene BCSG1, which was expressed in high abundance in a breast cancer cDNA library but scarcely in a normal breast cDNA library, was identified as a putative breast cancer marker. *In situ* hybridization analysis demonstrated a stage-specific BCSG1 expression: undetectable in normal or benign breast lesions, partial expression in ductal carcinoma *in situ*, but extremely high level in advanced infiltrating breast cancer. BCSG1 overexpression may indicate breast cancer malignant progression.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer, Invasion and Metastasis, Matrix Degradation, Metalloproteinases, Type IV Collagenases	69
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

✓ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

✓ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

✓ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

✓ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature

9/18/90
Date

TABLE OF CONTENTS

Adjustment

Metastatic dissemination is the primary cause of death for most breast cancer patients. The research effort in PI's lab is to uncover the mechanisms whereby breast cancer undergoes malignant progression and becomes metastatic. The onset and progression of breast cancer is accompanied by multiple genetic changes that result in qualitative and quantitative alterations in individual gene expression (1). Our hypothesis is that many of these quantitative genetic changes manifest themselves as alterations in the cellular complement of novel transcribed mRNAs. Identification of these mRNAs, if sufficiently characterized, could provide clinically useful information for patient management and prognosis while enhancing our understanding of breast cancer pathogenesis. Although pathological endpoints such as tumor size, lymph node status and status of estrogen receptor and progesterone receptor remain the most useful guides in prognosis and selecting treatment strategies for breast cancer (2), there is a need to further investigate the molecular mechanisms that determine the properties of an individual tumor e.g., probability of metastasis. While numerous prognostic factors have now been identified, few have contributed to defining clinical response to therapy.

The current Career Development Grant was initially awarded to study the novel 80kDa matrix degrading proteinase in breast cancer progression. During the past two years, we were not satisfied with the progress of this project. We have devoted many effort on raising monoclonal antibodies to the 80kDa proteinase in attempt to using these antibodies for purification and molecular cloning. Although we obtained the antiserum which can immunoprecipitate the 80kDa proteinase, no success has been met for development of monoclonal antibodies. Based on this unsatisfied work on the cloning of 80kDa proteinase, PI has made some adjustment on his effort.

With the availability of tens of thousands of partial cDNA sequences (EST: expressed sequence tag), researchers now shift their attention to the unveiling of expression profile of individual genes or pattens of genes in normal versus diseased states. Several newly developed strategies, such as Serial Analysis of Gene Expression (SAGE) (3) and cDNA Microarray method (4), have demonstrated potential for broad application for quantitative analysis of differential patterns of gene expression. Within this context, we undertook a search, using the differential cDNA sequencing approach (described in the manuscript 2), for isolation of differentially expressed sequence tags and the possible presence of the new marker genes for breast cancer.

Within the same research area of breast cancer metastasis, we recently identified and cloned two novel genes: tissue inhibitor of metalloproteinases-4, TIMP-4, and a putative breast cancer specific gene, BCSG1. The expressions of both TIMP-4 and BCSG1 in human breast tissue, including normal reduction mammoplasty specimens, benign breast lesions, carcinoma *in situ*, and infiltrating breast carcinoma, were examined. In addition, we also demonstrated an inhibitory effect of TIMP-4 on breast cancer growth and metastasis in the nude mice model. Currently, PI has one TIMP-4 paper in press in J. Biol. Chem; one TIMP-4 manuscript submitted to Cancer Res., and one BCSG1 manuscript submitted to Cancer Res. (see attached copies).

Project 1: TIMP-4, a novel tissue inhibitor of metalloproteinase whose expression is lost during breast cancer progression, inhibits growth of human breast cancers in the mammary fat pads of nude mice.

Introduction. A novel tissue inhibitor of metalloproteinases, TIMP-4, was cloned in PI's lab. Its expression in a variety normal tissue and in human breast cancer cells was examined. The anti-MMP activity of TIMP-4 was also confirmed in the conditioned medium of TIMP-4 transfected human breast cancer cells (please see attached manuscript 1).

Expression of TIMP-4 mRNA in human breast cancer cells. In order to select a suitable breast cancer cell line for TIMP-4 mediated gene transfection, the expression of TIMP-4 in human breast cancer cells was first investigated. As demonstrated in Fig. 1, Northern blot analysis failed to detect the TIMP-4 transcript in most breast cancer cell lines except MDA-MD-231 cells which showed a TIMP-4 transcript of 1.4 kb. The inability to pick up the TIMP-4 mRNA in most breast cancer cell lines by Northern blot suggests 1) that the TIMP-4 gene may be only expressed very weakly in breast epithelial cells but is mainly expressed in stromal cells, or 2) the expression of the TIMP-4 gene may be down-regulated in breast cancers during the breast malignant progression.

Transfection and selection of TIMP-4 positive clones. We selected MDA-MB-435 cell line as receipt for TIMP-4 mediated gene transfection because of 1) its lack of detectable of TIMP-4 transcript; and 2) its more aggressive and highly tumorigenic behavior in nude mice. The full-length TIMP-4 cDNA was inserted into pCI-neo mammalian expression vector; and the resulting vector was transfected into MDA-MB-435 cells. The same cells were also transfected with the vector containing no insert, as a control for TIMP-4. Transfection was repeated once, with different amounts of DNA, to allow establishment of neomycin-resistant clonal cell lines from independent transfections. Subsequent to transfection, G418 selection, and cloning by limiting dilution, several subclones of MDA-MB-435 cell were obtained. MDA-MB-435 subclones transfected with TIMP-4 cDNA were designated TIMP4-MDA-435, and MDA-MB-435 subclones transfected with pCVneo were designated neo-MDA-435. These G418-resistant clones were expanded into individual cell lines and used as a source for RNA and protein analysis. Clones were initially screened by *in situ* hybridization with a specific TIMP-4 antisense probe, and the positive clones were subjected to Northern blot analysis. Eight TIMP4-MDA-435 clones were picked up by *in situ* hybridization (Data not shown), and three clones were found to express a single mRNA band consistent with the size of the 1.4 kb TIMP-4 transcript (ref 5) (see Fig. 5A in manuscript 1). In contrast, none of 3 neo-MDA-435 clones produced any detectable TIMP-4 transcripts. Two high TIMP-4 expressing clones, TIMP4-MDA-435-19 and TIMP4-MDA-435-20, and two TIMP-4 negative (neo only) clones, neo-MDA-435-15 and neo-MDA-435-12, and parental MDA-MB-435 cells were chosen for further study. No changes in morphology were observed in these clones.

Expression of MMP inhibitory activity. The anti-MMP activity of TIMP-4 transfected clones was characterized. Conditioned media (CM) from two TIMP-4 positive clones (TIMP4-

MDA-435-19, TIMP4-MDA-435-20), and one TIMP-4 negative clones (neo-MDA-435-12) were collected, concentrated, and analyzed for metalloproteinase inhibitory activity by reverse zymography. Fig. 5C in manuscript 2 shows that the CMs from TIMP-4-producing clones contained a prominent MMP inhibitory activity at 22 kDa band in a non-reducing gelatin containing SDS gel. In contrast, no such activity was observed in the CM from neo-MDA-435 cells. These data suggest: 1) the TIMP-4 positive clones secret a functional TIMP-4-mediated anti-MMP activity; 2) no endogenous TIMP activities were detectable in neo-MDA-435 clones in the same conditions for detection of recombinant TIMP-4 activity.

In vitro growth of TIMP4-MDA-435 cells. To determine whether TIMP-4 expression affects the growth of MDA-MB-435 cells, the growth curve of TIMP4-MDA-435 cells was compared to that of neo-MDA-435 cell in monolayer culture. There was no significant difference in the growth pattern between parental MDA-MB-435 cells, neo-MDA-435 cells, and TIMP4-MDA-435 cells (data not shown).

Effect of TIMP-4 transfection on tumorigenicity. The tumorigenicity of TIMP4-MDA-435 cells was determined in comparison with parental MDA-MB-435 cells and neo-MDA-435 cells by inoculating $3x10^5$ cells into mammary fat pad of female nude mice. The growth of developing tumors was measured subsequently at regular intervals for six weeks. Three independent experiments were done to confirm reproducibility, and the data from three experiments are summarized in Table 1. After a lag phase of 7-10 days, mice given implants of both TIMP-4 positive and TIMP-4 negative cells developed tumors. There was no difference in tumor incidence among the groups. As demonstrated in Fig. 2, after a slow growth phase of 17 days, tumors from parental MDA-MB-435 cells increased in volume at an exponential rate. Starting at about 25 days after inoculation, great level of tumor narcosis was observed in tumors derived from MDA-MB-435 cells. The same breast cancer cells transfected with TIMP-4, however, were significantly inhibited in their tumor growth in vivo; and no tumor narcosis was observed. The mean volume of TIMP4-MDA-435-20 tumor was only 7% of that in parental MDA-MB-435 cells, 37% of that in neo-MDA-435-15 cells, and 22% of that in neo-MDA-435-12 cells ($P<0.01$ by two sided Student's t test). Two of three TIMP-4 positive clones, TIMP4-MDA-435-20 and TIMP4-MDA-435-4, showed a decreased tumor growth compared with parental MDA-MB-435 cells and neo-MDA-435 cells (all $p < 0.01$; Table 1). One of the TIMP-4 positive clone, TIMP4-MDA-435-12, exerted the similar tumor growth rate in nude mice compared with TIMP-4 negative clones. This lack of the inhibitory effect is due to the loss of TIMP-4 expression in TIMP4-MDA-435-4 cells in the *in vivo* environment (data not shown). Thus, the tumorigenicity of the breast cancer cells was inhibited by expression of TIMP-4.

Summary. Recently, we identified, cloned, and characterized a novel human tissue inhibitor of metalloproteinases-4, TIMP-4. To determine if TIMP-4 can modulate the *in vivo* growth of human breast cancers, we transfected a full-length TIMP-4 cDNA into MDA-MB-435 human breast cancer cells and studied the orthotopic growth of TIMP-4-transfected (TIMP4-MDA-435) vs control (neo-MDA-435) clones in the mammary fat of athymic nude mice. TIMP4-MDA-435 clones expressed TIMP-4 mRNA and produced an anti-metalloproteinase (MMP) activity detected by reverse zymography; while neo-MDA-435 clones did not express TIMP-4 mRNA or produce detectable anti-MMP activity. in vitro, TIMP4-MDA-435 clones showed no

significant difference in cell proliferation as compared with controls. When these cells were injected orthotopically in nude mice, we found that the overexpression of TIMP-4 significantly inhibited tumor growth rates; reached (4-10)-fold smaller primary tumor volumes at sacrifice ($p < 0.01$); and gave lower rates of axillary lymph node and lung metastasis, as compared with neo-MDA-435 clones.

Project 2: identification of a breast cancer specific gene, BCSG1, by direct differential cDNA sequencing

Introduction. Identification of quantitative changes in gene expression that occur in the malignant mammary gland, if sufficiently characterized, may yield novel molecular markers which may be useful in the diagnosis and treatment of human breast cancer. Several differential cloning methods, such as differential display polymerase chain reaction and subtractive hybridization, have been used to identify the genes differentially expressed in breast cancer biopsies, as compared to normal breast tissue controls (6-10). However, these investigations have involved the relatively time- and labor-intensive steps of subcloning, library screening, and cDNA sequencing of individual genes (7,11). On the other hand, creation of expressed sequence tag libraries is a rapid method used to identify or "tag" sequences that are expressed in specific tissues (12-13). Since the introduction of the EST sequencing approach, many novel human genes have been discovered (12-13). The advantage of this methodology, compared to isolation and sequencing of individual cDNAs, is that a large number of sequences can be "cataloged" with small amounts of sequencing data.

In this initial report, we described a novel breast cancer specific gene named BCSG1 that is overexpressed in advanced infiltrating breast cancer cells but not in normal or benign breast lesion. The expression pattern of BCSG1 may be a meaningful marker in the development of breast cancer. Please see attached manuscript 2 for the table and figures.

Molecular cloning of BCSG1. BCSG1 was identified and cloned by differential cDNA sequencing as described in the manuscript 2. Comparison of the predicted amino acid sequence with the sequence of a similar human protein was analyzed. After optimal alignment, the putative BCSG1-encoded protein shows 54% sequence identity with the recently cloned non-Aß fragment of human Alzheimer's disease (AD) amyloid protein (14).

Tissue expression. The expression of BCSG1 gene in a variety of normal human tissues were analyzed by Northern blotting (Fig. 3 in manuscript 2). As expected, the Northern blot showed that BCSG1 was abundantly expressed as a 1 kb transcript in brain which is the rich source for AD amyloid family gene. Similar bands with much lower accumulations in their relative intensity were also obtained in ovary, testis, colon, and heart. By contrast, none of them was present in other specimens analyzed, such as breast, kidney, liver, prostate, lung, small intestine, thymus and placenta.

Expression of BCSG1 in human breast cancer cells. In an attempt to evaluate the potential biological significance of BCSG1 on human breast cancer development and progression, we studied BCSG1 gene expression in human breast biopsy samples. The expression of BCSG1 in metastatic breast carcinoma and normal breast tissue were analyzed by Northern blotting. Fig.

4 in manuscript 2 shows overexpression of BCSG1 transcript in an infiltrating breast carcinoma. In contrast, the BCSG1 transcript was undetectable in normal breast tissue. The presence of BCSG1 transcript in human breast tissue and its overexpression in breast carcinomas are consistent with our differential cDNA sequencing cloning strategy which suggests a possible role or a biomarker of up-regulation of BCSG1 in the development of breast cancer.

The expression of BCSG1 was also investigated in a variety of human breast cancer cell lines (Fig. 5 in manuscript 2). Northern blot detected the 1 Kb BCSG1 transcript in 2/4 lines derived from pleural effusion and 4/4 lines detected from ductal infiltrating carcinomas. Among these lines, H3922 expressed the highest level of BCSG1 mRNA. The absence of BCSG1 mRNA in some breast cancer cell lines may suggest that the expression of BCSG1 gene requires specific in vivo conditions or that it is induced by interactions between the tumor cells and stromal cells.

In order to localize the cellular source of the BCSG1 expression and to further assess the biological relevance of the overexpression of BCSG1 in breast cancers, we next performed *in situ* hybridization on fixed breast sections from 20 infiltrating carcinomas, 15 ductal carcinomas *in situ* (DCIS), and 18 benign breast lesions including 5 reduction mammoplasty specimens, 8 breast hyperplasias, and 5 fibroadenomas. In these experiments, we examined two aspects of BCSG1 expression: 1) the tissue localization (stromal versus epithelial); and 2) the correlation of BCSG1 expression and breast cancer malignant phenotype. There was a wide variation in staining intensity for BCSG1 expression among the human breast cancer specimens. Since the colorimetric *in situ* hybridization is not quantitative, the tissue samples were classified into either positive or negative staining for BCSG1 expression; no attempt was made to differentiate the levels of expression of BCSG1 among positive-staining specimens. The negative cases were confirmed with at least two independent experiments. All stainings were reviewed by at least two people. Fig. 6 in manuscript 2 shows a representative *in situ* hybridization for BCSG1. We found a strongly positive BCSG1 hybridization in neoplastic epithelial cells of highly infiltrating breast carcinomas (Fig. 6A,B). The expression of BCSG1 mRNA was detectable in the neoplastic epithelial cells in 17 of 20 infiltrating breast carcinomas. No expression of BCSG1 was detected in the stromal cells. In contrast, expression of BCSG1 was absent in 16 out of 18 cases of normal or benign breast lesions. A representative negative staining of BCSG1 in normal ductal breast epithelial cells (Fig. 6E), a benign proliferative breast lesion (Fig. 6F), and a benign fibroadenoma (Fig. 6G) are presented. Furthermore, as demonstrated in Fig. 6B for a highly invasive breast carcinoma, no detectable signal of BCSG1 expression was evident in the residual normal lobular breast epithelial cells although the surrounding invasive breast carcinoma cells were stained positive for BCSG1 expression. These *in situ* hybridization results are consistent with the Northern blot analysis which showed a strong expression of BCSG1 transcript in breast carcinoma but not in normal or benign breast lesions.

It is interesting to note that although a strong BCSG1 signal was easily detected in the malignant breast epithelial cells of infiltrating breast carcinoma, the *in situ* carcinomas showed a different BCSG1 expression patterns. Among 15 DCISs (8 are Comedo type and 7 are non-Comedo type), 8 specimens were stained negatively (Fig. 6D) and 7 specimens were positive (Fig. 6C). Interestingly, 6 of 7 BCSG1 positive DCIS samples were Comedo type DCIS and only one was non-Comedo type; among the BCSG1 negative specimens, there were 6 non-Comedo type DCISs and only two Comedo type DCISs. These results, which demonstrated a stage-specific BCSG1 expression from virtually no detectable expression in normal or benign breast to partial expression (7/15) in the *in situ* breast carcinoma and to the high expression (17/20) in the infiltrating malignant

breast carcinomas, suggest an association of BCSG1 expression with breast cancer malignant progression. Based on this BCSG1 expression pattern, we propose that BCSG1 may be potentially used as a breast cancer progression marker.

Summary. With the availability of tens of thousands of partial cDNA sequences, we have, using differential cDNA sequence, identified a new putative breast cancer marker gene BCSG1 and studied its expression in breast cancer. Using *in situ* hybridization analysis, we have demonstrated the expression of BCSG1 transcripts in the neoplastic epithelial cells of infiltrating breast carcinoma but not in epithelial cells of normal and benign breast. The overexpression (13 of 15) of BCSG1 in malignant infiltrating breast epithelial cells compared to the partial expression (7 of 15) in the *in situ* carcinoma suggests that up-regulation of BCSG1 expression is associated with breast malignant progression and may signal the more advanced invasive/metastatic phenotype of human breast cancer. This implication is further supported by detection of BCSG1 expression in 6/8 aggressive Comedo type DCISs and only 1/7 in non-Comedo type DCISs. It is unlikely that BCSG1 is overexpressed as a secondary effect of cellular proliferation because no detectable BCSG1 expression is evident in rapidly proliferating nonmalignant breast lesions.

It is interesting to note that the predicted amino acid sequence of BCSG1 gene shares high sequence homology with the recently cloned non-Aß component of Alzheimer's disease (AD) amyloid precursor protein (14). A neuropathological hallmark of AD is a widespread amyloid deposition resulting from beta-amyloid precursor proteins (beta APPs). Beta AAPs are large membrane-spanning proteins that either give rise to the beta A4 peptide (Aß fragment) (15) or a non-Aß component of AD amyloid (14) that is either deposited in AD amyloid plaques or yielding soluble forms. While the insoluble membrane-bound AD amyloid destabilizes calcium homeostasis and thus renders cell vulnerable to excitotoxic conditions of calcium influx resulting from energy deprivation or overexcitation (16), the soluble AD amyloid proteins are neuroprotective against glucose deprivation and glutamate toxicity, perhaps through their ability to lower the intraneuronal calcium concentration (17). We currently do not know whether BCSG1 is an instigator or merely a by-product during breast cancer progression. With the availability of anti-BCSG1 antibody to localize BCSG1 protein and the recombinant BCSG1 protein, we may start to speculate that BCSG1, like soluble AD amyloid, may be potentially involved in tissue damage resulting from tissue remodeling due to the local cancer invasion. An elucidation of the reasons for BCSG1 overexpression in infiltrating breast cancer cells may shed some light on the pathogenesis of breast cancer progression. Nevertheless, we demonstrated a stage-specific BCSG1 expression and an association of BCSG1 overexpression with clinical aggressiveness of breast cancers. The notion that the BCSG1 overexpression may indicate breast cancer malignant progression from benign breast or low grade *in situ* carcinoma to the highly infiltrating carcinoma warrants further investigation.

References:

1. Porter-Jordan, K. and Lippman, M.E. Overview of the biological markers of breast cancer. *Hematol. Oncol. Clin. N. Am.*, **8**: 73-100, 1994.
2. Manning, D.L., McClelland, R.A., Knowlden, J.M., Bryant, S., Green, C.D., Sutherland, R.L. and Ormandy, C.J. Differential expression of oestrogen regulated genes in breast cancer. *Acta Oncol.*, **34**: 641-646, 1995.
3. Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. Serial analysis of gene expression. *Science*, **270**: 484-7, 1995.
4. Schena, M., Shalon, D.,Davis, R.W. and Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**: 467-70, 1995.
5. Damon A. Douglas, Yuenian Eric Shi and Qingxiang Army Sang. Computational sequence analysis of the tissue inhibitor of metalloproteinase family, *J. of Protein Chem.*, in press.
6. Watson, M.A. and Fleming, T.P. Isolation of differentially expressed sequence tags from human breast cancer. *Cancer Res.*, **54**: 4598-4602, 1994.
7. Sager, R., Anisowicz, A., Neveu, M., Liang, P. and Sotiropoulou, G. Identification by differential display of alpha 6 integrin as a candidate tumor suppressor gene. *FASEB J.*, **7**: 964-970, 1993.
8. Chen, Z. and Sager, R. Differential expression of human tissue factor in normal mammary epithelial cells and in carcinomas. *Mol. Med.*, **1**: 153-160, 1995.
9. Zhang, M., Zou, Z., Maass, N. and Sager, R. Differential expression of elafin in human normal mammary epithelial cells and carcinomas is regulated at the transcriptional level. *Cancer Res.*, **55**: 2537-2541, 1995.
10. Zou, Z., Anisowicz, A., Hendrix, M.J., Thor, A., Neveu, M., Sheng, S., Rafidi, K., Seftor, E. and Sager R. Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. *Science*, **263**: 526-529.
11. Liang, P., Averboukh, L., Keyomarsi, K., Sager, R. and Pardee, A.B. Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. *Cancer Res.*, **52**: 6966-6968, 1992.
12. Adams, M.D., Soares, M.B., Kerlavage, A.R., Field, C. and Venter, J.C. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.*, **4**: 373-380, 1993.
13. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Oide, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., and Venter, J.C. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science.*, **252**: 1651-6, 1991.
14. Ueda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A, Yoshimoto, M., Otero, D.A., Kondo, J., Ihara, Y. and Saitoh, T. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. U.S.A.*, **90(23)**: 11282-6, 1993.
15. Masters, C.L., Simms, G., Weinman, N.A., Multhaup, G., McDonald, B.L. and Beyreuther, K. Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc. Natl. Acad. Sci. USA* **82**: 4245-4249, 1985.
16. Mattson, M.P., Rydel, R.E., Lieberburg, I., Smith-Swintosky, V.L. Altered calcium signaling and neuronal injury: stroke and Alzheimer's disease as example. *Ann. N. Y. Acad. Sci.*, **679**:

1-*21*, 1993.

17. Barger, S.W., Fiscus, R.P., Ruth, P., Hofmann, F. and Mattson, M.P. Role of cyclic GMP in the regulation of neuronal calcium and survival by secreted forms of beta-amyloid precursor. *J. Neurochem*. **64**: 2087-96, 1995.

Table 2. Primary tumor size, lymph node status, and lung metastases at sacrifice

Experiment	Treatment Group	Tumor Vol (mm^3) of Primary Size	Tumor Incidence Tumor/Total (%)	Lymph Nodes Pos/Total (%)	Lung Metastases Pos/Total (%)
1	TIMP4-MDA-435-20	580 ± 223	9/10 (90%)	0/8 (0)	0/5 (0)
	TIMP4-MDA-435-12	2200 ± 819	9/10 (90)	2/10 (20)	2/5 (40)
	Neo-MDA-435-12	2550 ± 404	10/10 (100)	4/10 (40)	2/5 (40)
2	TIMP4-MDA-435-4	860 ± 165	8/10 (80)	-	-
	TIMP4-MDA-435-20	770 ± 230	7/10 (90)	-	-
	Neo-MDA-435-15	2060 ± 95	10/10 (100)	-	-
3	TIMP4-MDA-435-20	254 ± 86	7/10 (70)	0/9 (0)	0/5 (0)
	MDA-MB-435	1684 ± 351	8/10 (80)	5/10 (50)	3/5 (60)

For each line tested, 2 x 10^5 cells were implanted into the mammary fat pad; 2 injections per mouse. Each experimental group consisted of 5 animals, which were given independent injections. All animals were sacrified at 35 days after cell implantion. Tumor volumes are expressed as means ± SEMs (number of tumors assayed). Statistical comparisions of pooled data for TIMP4-MDA-435 tumor vs neo-MDA-435 or MDA-MB-435 tumors: 1) TIMP4-MDA-435-20 vs neo-MDA-435-12: $p < 0.01$; 2) TIMP4-MDA-435-4 and TIMP4-MDA-435-20 vs neo-MDA-43515: $p < 0.01$; TIMP4-MDA-435-20 vs MDA-MB-435: $p < 0.001$.

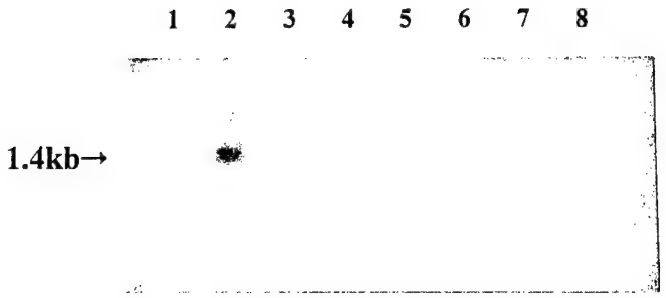


Fig. 1. Northern analysis of TIMP-4 expression in human breast cancer cells. RNAs were isolated and subjected to Northern analysis by hybridization with random-labelled full-length cDNA probe. The integrity of the RNA samples was ascertained by direct visualization of the ribosomal RNAs in stained gel. Each lane contained 20 μg total RNA. Northern analysis failed to detect the TIMP-4 transcript in most breast cancer cell lines, except MDA-MB-231 cells, which showed a strong 1.4 kb TIMP-4 transcript. 1. Hs578t; 2. MDA-MB-231; 3. MDA-MB-435; 4. MDA-MB-436; 5. MCF-7; 6. T47D; 7. BT549; 8. TKS-7 (FGF-4 transfected T47D cells).

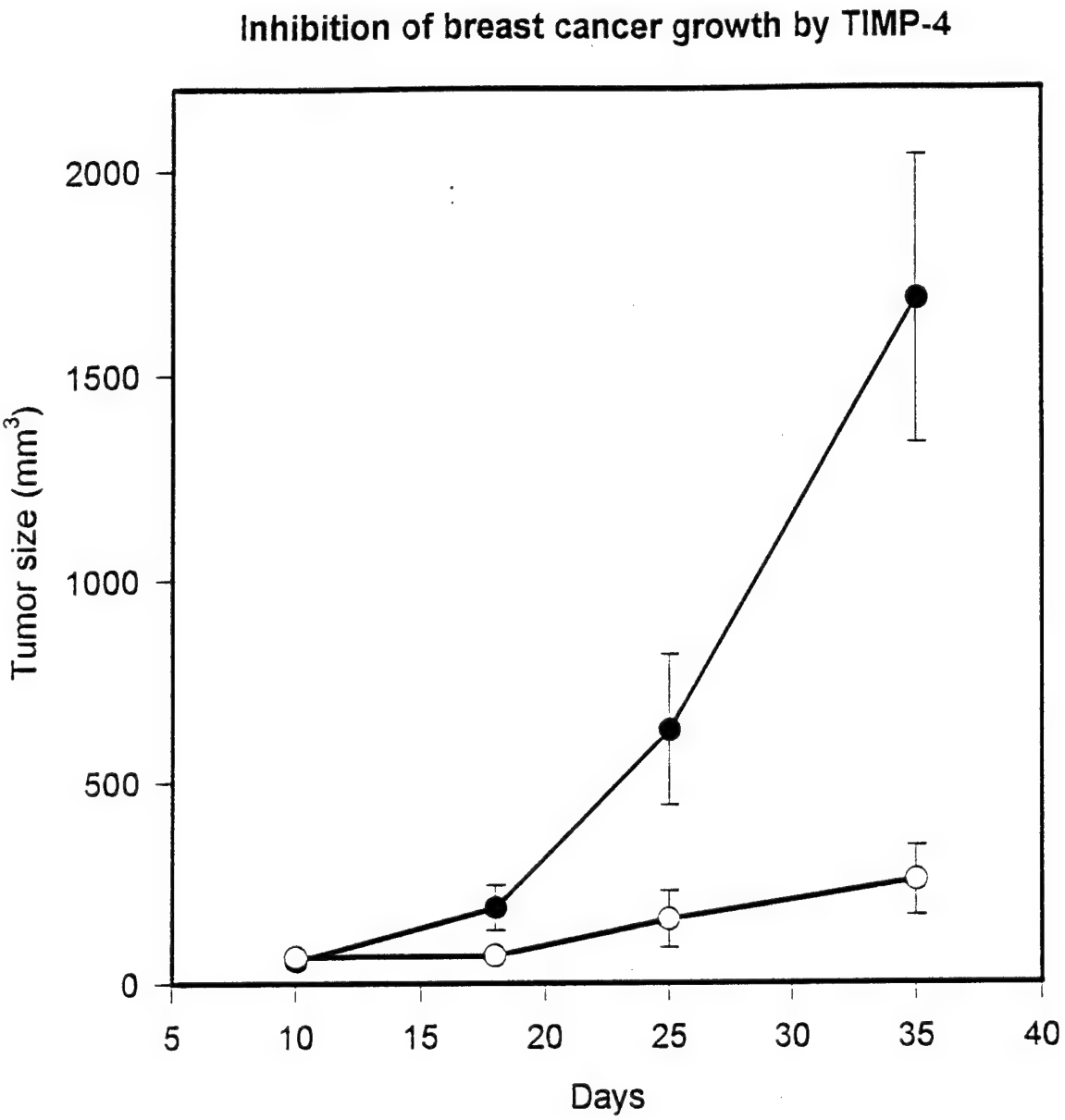


Fig. 2. Inhibition of tumor growth in nude mice. Closed circle represents parental MDA-MB-435 cells and open circle represents TIMP4-MDA-435-20 cells. The data reflect the experiment 3 in Table 1. Cell injection and tumor measurement were described in Table 1.

The Journal of Biological Chemistry
Published by
American Society for Biochemistry and Molecular Biology

ROBERT T. SCHIMKE
Associate Editor

Department of Biological Sciences
Stanford University
Stanford, CA 94305-5020
Tel: (415) 725-4814
Fax: (415) 723-0155

FAX TRANSMITTAL FORM

TO: Dr. Yuenian E. Shi

DATE: 8/23/96

Phone: 718-470-3086

FAX: 718-470-6744

ADDRESS: Pediatric Research Center
Schneider Children's Hospital
Long Island Jewish Medical Center
New Hyde Park, NY 11042

Your manuscript(s) M6-2554

"Molecular Cloning and Characterization of Human Tissue Inhibitor ..."

is (are) accepted for publication and is (are) tentatively scheduled for a Oct./Nov. issue. You will receive page proof; this should be returned promptly to expedite publication. You will be billed after publication for page charges, reprints you order, half-tones, electron micrographs and authors' alterations as applicable.

Robert T. Schimke, Associate Editor

Molecular cloning and characterization of human tissue inhibitor of metalloproteinases-4 (TIMP-4)

John Greene[1], Mingsheng Wang[2], Lisa A. Raymond[1],
Yiliang E. Liu[2], Craig Rosen[1] and Yuenian E. Shi[2]

[1]Human Genome Sciences, Inc., Rockville, MD 20850-3338.

[2]Department of Pediatrics
Long Island Jewish Medical Center,
The Long Island Campus for
Albert Einstein College of Medicine
New Hyde Park, NY 11042

Corresponding author
Dr. Yuenian Eric Shi
Pediatric Research Center
Schneider Children Hospital
Long Island Jewish Medical Center
New Hyde Park, NY 11042
Phone 718-470-3086
FAX 718-470-6744
e-mail Shi@aecom.yu.edu

Running title: A novel human tissue inhibitor of metalloproteinases, TIMP-4

Abbreviations: MMP, matrix metalloproteinase; TIMP, tissue inhibitor of metalloproteinases; ECM, extracellular matrix; EST, expressed sequence tag

Key words: metalloproteinase, TIMP

Abstract

The tissue inhibitors of metalloproteinases (TIMPs) comprise a family of proteins, of which three members have so far been described. Using the expressed sequence tag (EST) sequencing approach, we have identified a novel TIMP related cDNA fragment and subsequently cloned a fourth human TIMP (TIMP-4) from human heart cDNA library. The open reading frame (ORF) encodes a 224 amino acid precursor including a 29-residue secretion signal. The predicted structure of the new protein shares 37% sequence identity with TIMP-1 and 51% identity with TIMP-2 and TIMP-3. The protein has a predicted isoelectric point of 7.34. The ORF directed expression of TIMP-4 protein in MDA-MB-435 human breast cancer cells showed metalloprotease inhibitory activity on reverse zymography. By Northern analysis, only adult heart showed abundant TIMP-4 transcripts with a 1.4 kb predominant transcript band; very low levels of the transcripts were detected in kidney, placenta, colon and testes; and no transcripts were detected in liver, brain, lung, thymus and spleen. This unique expression pattern suggests that TIMP-4 may function in a tissue-specific fashion in extracellular matrix (ECM) homeostasis.

Introduction

Matrix metalloproteinases (MMPs) play a critical role in ECM homeostasis. Controlled remodeling of the ECM is an essential aspect in the process of normal development, and deregulated remodeling has been indicated to have a role in the etiology of diseases such as arthritis, periodontal disease, and cancer metastasis (1-5). The overproduction and unrestrained activity of MMPs has been linked to malignant conversion of tumor cells [4-12]. The down-regulation of MMPs may occur at the levels of transcriptional regulation of the genes; activation of secreted proenzymes; and through interaction with specific inhibitor proteins, such as TIMPs. TIMPs are secreted multifunctional proteins that play pivotal roles in the regulation of ECM metabolism. Their most widely recognized action is as inhibitors of matrix MMPs. Thus, the net MMP activity in the ECM is the result of the balance between activated enzyme levels and TIMPs levels. Augmented MMP activity is associated with the metastatic phenotype of carcinomas, especially breast cancer [7-9, 13-16]; the decreased production of TIMP could also result in greater effective enzyme activity and invasive potentials [17-19]. These results suggest that an increase in the amount of TIMPs relative to MMPs could function to block tumor cell invasion and metastasis. In fact, tumor cell invasion and metastasis can be inhibited by up-regulation of TIMP expression or by an exogenous supply of TIMPs [17, 40-44].

Three mammalian TIMPs have been characterized at the sequence level: TIMP-1 (20), TIMP-2 (21) and TIMP-3 (22-25, 35,54). The proteins are classified based on structural similarity to each other, as well as their ability to inhibit matrix metalloproteases. There have been other reports of inhibitors of metalloproteases (IMPs) with characteristics different from these known TIMPs. In some cases these activities result from alternate forms of the known TIMPs. For

instance, a report describes one IMP present in the conditioned media of human bladder carcinoma to be a partially glycosylated form of TIMP-1 and another to be a partially processed/degraded form of TIMP-2 (25). There are additional reports that describe sources and characteristics of IMP activity, but the gene products associated with these activities have not been delineated (26).

Individual TIMP family members may have specific physiological roles. This notion is supported by several lines of evidence. First, although TIMPs are essentially interchangeable in their capabilities as inhibitors of MMPs, they are distinguished by the formation of specific complexes with different pro-MMPs (27-29). Secreted MMP-2/TIMP-2 and MMP-9/TIMP-1 complexes may represent an additional function for TIMPs in controlling activation of specific latent MMPs. Unlike TIMP-1 and TIMP-2, TIMP-3 has a unique association with ECM (30). Second, the expression of TIMP genes is quite different. The TIMP-1 gene is highly inducible at the transcriptional level in response to many cytokines and hormones (31-34). Likewise, TIMP-3 expression is not only induced in response to mitogenic stimulation, but also is subject to cell cycle regulation (35), suggesting that TIMP-3 expression may represent an invaluable tool for the analysis of cell cycle progression, terminal differentiation, and replicative senescence. In contrast, TIMP-2 expression, like that of MMP-2 with which it interacts, is largely constitutive (36-37).

Since the introduction of the expressed sequence tag (EST) sequencing approach, many novel human genes have been discovered and isolated [38]. With the rapidly growing repertoire of human ESTs, we took advantage of automated EST sequence analysis to identify novel TIMP-related genes. We have described here the full-length sequence of a novel member of the TIMP family and examined the expression of this new member, TIMP-4, in a variety of tissues. We

have also demonstrated an MMP inhibitory activity of the expressed TIMP-4 protein.

Materials and Methods

Reagents. Restriction enzymes, T7 polymerase, random primer DNA labeling kit, and digoxigenin-labeled nucleotides were obtained from Boehringer Mannhem, Indianapolis. ^{32}P-dATP was purchased from Amersham.

Molecular cloning of TIMP-4 full-length cDNA sequence. We have used EST analysis to search for a new TIMP. A data base containing approximately 500,000 human partial cDNA sequences (expressed sequence tags) has been established in a collaborative effort between the Institute for Genomic Research and Human Genome Science Inc., using high throughput automated DNA sequence analysis of randomly selected human cDNA clones (38). Sequences of TIMP-related genes were searched for using the blastn and tblastn algorithms [39]. An EST from a human brain library, which demonstrated homology to TIMPs, was completely sequenced and found to be a partial clone lacking the sequence at the 5' end. The coding region and 3' untranslated region of this clone was excised from the Bluescript vector by digestion with the restriction endonucleases EcoRI and XhoI, and used to generate a radiolabelled probe. This probe was used to screen a Northern blot of total RNAs from several human tissues. The highest level of expression of the putative novel TIMP was noted in RNA from adult heart. We next generated a cDNA library from human heart. Poly A mRNA from heart tissue was obtained using Oligotex beads. Five micrograms of this mRNA was used to construct a directional cDNA library in the Stratagene Unizap vector using the Stratagene cDNA library kit. One million clones of the primary library were amplified and an aliquot excised to yield Bluescript SK plasmid clones. These clones were

screened with the probe generated by EcoRI and XhoI digestion of the positive clone from human brain library as described above. Positive clones were re-screened, both by hybridization and PCR analysis, using a Bluescript reverse primer and an antisense primer (5' GACTGTCCACTTGGCACTTCT 3') specific for the putative TIMP-related gene in the 3' untranslated region. The full-length cDNA was completely sequenced using ABI 373a Automated Fluorescent Sequencer protocols.

Northern analysis. Total RNA was extracted from tissues according to the method of Chomcznski and Sacchi [45]. The RNA from human breast cancer cells was prepared using the RNA isolation kit RNAzol B (Tel-Test, Inc) based on the manufacturer's instruction. Equal aliquots of RNA were electrophoresed in a 1.2% agarose gel containing formaldehyde and transferred to nylon membrane (Boehringer Mannheim). The membrane was pre-hybridized with ExpressHyb hybridization solution (Clontech, Inc.) at 68^0C for 30 min. The hybridization was carried out in the same solution with ^{32}P-labeled TIMP-4 probe (1.5×10^6 cpm/ml) for 1 hour at 68^0C. The membrane was then rinsed in 2 x SSC containing 0.05% SDS three times for 30 min at room temperature, followed by two washes with 0.1 x SSC containing 0.1% SDS for 40 min at 50^0C. The full-length TIMP-4 cDNA was isolated from the Bluescript vector, following EcoRI and XhoI digestion, and used as a template for preparation of a random-labelled cDNA probe. The riboprobe is a 390 base segment extending from nucleotides 800 to 1,189 (the end of the 3'-end of the cDNA). This riboprobe, which covers 85% of the 3' untranslated region, was generated by PstI digestion of the Bluscript vector, followed by RNA synthesis with T7 polymerase.

Expression of TIMP-4 in human breast cancer cells. Human TIMP-4 full-length sequence was subcloned into the pCI-neo mammalian Expression Vector (Promega) downstream of the human

cytomegalovirus promoter to generate the pCITIMP4 expression vector. Forty micrograms of pCITIMP4 or the control pCI-neo plasmids were transfected into MDA-MB-435 human breast cancer cells by the calcium phosphate-mediated method as previously described (46). Thirty G418-resistant individual clones were selected in the selection medium containing 800 ug/ml of G-418, subcloned and characterized by *in situ* hybridization and Northern blot analysis. TIMP-4-producing clones were grown in serum-free defined medium. The conditioned media were collected at 40 hours after culturing cells in serum-free DMEM medium, concentrated approximately 10-fold using an Amicon hollow fiber concentrator with 10,000 molecular weight cut off. The inhibitory activity was subsequently analyzed on reverse zymography SDS-PAGE.

Electrophoretic analysis by reverse zymography. Samples of conditioned media from TIMP-4 -producing clones and control clones were adjusted to the same protein concentration and electrophoresed on a 0.1% SDS, 12% polyacrylamide protease/substrate gel (47). The gel was incubated in the collagenase buffer (50 mM Tris, pH 7.4, 0.2 M NaCl, 5 mM $CaCl_2$, 1% Triton X-100, 0.7 ug/ml of recombinant MMP2) at 37^{0}C overnight to allow digestion of gelatin in the gel. The MMP inhibitory activities of samples were visualized by commassie blue R-250 (Sigma) staining and destaining as previously described (48).

Results

Molecular cloning of TIMP4 complementary DNA. We have searched a database of human genes identified by the EST method. The automated screening revealed an EST from a human brain library with a 45% sequence homology to TIMP-2 protein. The clone was completely sequenced. A putative stop codon was located; however, a start codon (ATG) could not be located at the 5' end. The length of the open reading frame was also shorter than expected for a 22-28 kDa protein in TIMP family. Therefore, it was concluded that this cDNA clone did not encode the entire putative TIMP protein and that a segment at the 5' end containing the start codon was missing. In order to obtain the full-length sequence of the putative new TIMP gene, the identified cDNA clone was prepared as a probe and was used to investigate the expression of this new putative TIMP gene in a variety of human tissues by Northern blot analysis. Because the highest expression of this new putative TIMP gene was identified in human heart, we next generated a cDNA phage library from human adult heart and screened one million clones for additional 5' sequence. As result, a number of clones were identified and the longest of these was sequenced and found to contain the full-length cDNA sequence of the putative new TIMP gene.

The nucleotide sequence determined from this clone and the predicted corresponding amino acid sequence are shown in Fig. 1. The full-length cDNA sequence contains 1,189 bp with 672 bp open reading frame; 59 bp in the 5' untranslated region; and 458 bp of 3' untranslated sequence. The open reading frame extends from the initiation $A_{60}TG$ codon to TAG_{732} stop. The open reading frame encodes a protein of 224 amino acids. A hydrophobic leader sequence at the amino terminus conforms to a consensus signal peptide with a predicted cleavage site following an alanine residue located at position 29 in the precursor (Fig. 1). Removal of the signal sequence

results in a mature protein of 195 amino acids having a calculated molecular weight of 22 kDa, which is in close agreement with the molecular mass range of the TIMP family. The deduced amino acid sequence predicts a protein with an isoelectric point of 7.34. Comparison of the predicted amino acid sequence with the sequences of human TIMP-related proteins is shown in Fig. 2. After optimal alignment, the putative protein shows 37% sequence identity and 57% similarity to TIMP-1 and 51% identity and 70% similarity to TIMP-2 and TIMP-3. These calculations do not take into account the significance of any gaps in the alignments. The predicted protein structure of the putative new protein shares several essential features that are characteristic to the TIMP family, including 12 completely conserved cysteine residues in the corresponding positions that form intrachain disulfide bonds that fold the protein into a two-domain structure [49]. The presence of a consensus sequence VIRAK, which has been proposed to serve a hallmark of the TIMP family (50,54), was also observed in the most conserved first 22 amino acids located at the N-terminal region.

The extensive similarity of the predicted amino acid sequence with TIMPs suggests that the putative new protein is a novel member of the human TIMP family and should be designated as human TIMP-4.

Tissue expression. Tissue-specific transcription of TIMP-4 was examined by Northern blotting on 20 micrograms of total RNAs from various human adult tissues (Fig. 3). As expected, the Northern blot showed maximal TIMP-4 transcript levels in heart. Using a full-length cDNA hybridization probe, transcripts of 4.1, 2.1, 1.4, 1.2 and 0.97 kb were detected in heart, with the 1.4-kb band representing at least 90% of the hybridization signal. Similar bands, with much lesser accumulations in their relative intensity, were also obtained in kidney, pancreas, colon and testes.

By contrast, none of them was present in other specimens analyzed such as liver, brain, lung, small intestine, thymus and spleen. The 1.4 kb TIMP-4 transcript was also detected in RNA isolated from the human breast cancer cell line MDA-MB-231 (Fig. 4). In order to rule out the possibility of cross-hybridization with either TIMP-1, TIMP-2, or TIMP-3; an additional filter with RNA from MDA-MB-231 cells was also hybridized with a 389 bp riboprobe, which represents a specific nucleotide sequence of the 3'-untranslated TIMP-4. As shown in Fig. 4B, the riboprobe recognized the same bands in the RNA from MDA-MB-231 cells as the complete DNA probe, thus suggesting that the 1.4 kb transcript corresponds to TIMP-4.

Expression of MMP inhibitory activity. Active recombinant TIMP-4 protein is required for characterization of its biochemical activity against MMPs and biological functions to inhibit tumor growth and metastasis. As an initial attempt to evaluate the biological significance of TIMP-4 to inhibit cancer growth and metastasis, we have transfected TIMP-4 full-length cDNA into the highly tumorigenic MDA-MB-435 human breast cancer cells. Three positive clones have been selected and expressed high levels of TIMP-4 transcript (Fig. 5A). Conditioned media (CM) from two TIMP-4 positive and one control clones were collected, concentrated, and analyzed for metalloproteinase inhibitory activity by reverse zymography. Fig. 5B shows that the CMs from TIMP-4-producing clones contained a prominent MMP inhibitory activity at 24 kDa band in a non-reducing gelatin containing SDS gel. In contrast, no such activity was observed in the CM form control MDA-MB-435 cells, suggesting that no endogenous TIMP activities were detectable in the same conditions for detection of recombinant TIMP-4 activity.

Discussion.

The work described here introduces a new member of the TIMP family, on which we confer the title TIMP-4 because of its high sequence homology to the TIMP family, 12 conserved cysteine residues, and the expressed MMP inhibitory activity.

The classical approach to identify novel proteins begins with the discovery of an interesting biological activity. This protein is then purified; biochemically characterized; and subsequently, the gene is cloned. Since the introduction of the EST sequencing approach and the availability of tens of thousands of ESTs, researchers can now shift their attention to high-throughput cDNA cloning in conjunction with structural similarity analysis as an accelerated method for protein discovery. In this regard, the nucleic acid sequences of randomly picked cDNAs from established EST data bases are searched and analyzed by the BLAST program for sequence similarity to the protein of interest. Where similarities are detected, it is possible to make functional inferences concerning the encoded protein based upon what is known about the function of the matched sequences. Using this approach, we identified an EST with high sequence homology to TIMP-2 and subsequently, the novel TIMP-4 gene was cloned using this EST as a probe.

The predicted protein structure of TIMP-4 shows several interesting features. First, as expected, essential features of other TIMPs are conserved, including the location of 12 Cys residues, as well as their relative spacing and the presence of 29-amino acid leader sequence, which presumably is cleaved to produce the mature protein (13). Second, the mature protein has an expected size of 22 kDa which is similar to the sizes of TIMP proteins. Expressed rTIMP-4 protein migrates as a 24 kDa protein by reverse zymography SDS-PAGE at non-reducing condition, which is consistent with that obtained for other TIMPs (55). Third, the deduced amino acid sequence of TIMP-4 predicts a protein with an isoelectric point of 7.34, the most neutral

human TIMP protein at the physiological condition (pH7.4) comparing to values of 8.00, 6.45, and 9.04 for human TIMP-1, TIMP-2, and TIMP-3, respectively (24). Fourth, as expected, TIMP-4 has a highly conserved N-terminal domain similar to other TIMPs. The N-terminal 126 amino acid residues of mature TIMP-1 (51) and the N-terminal 127 residues of mature TIMP-2 (52,53) have been shown to be adequate for the inhibition of MMPs, suggesting that this part of the proteins is functionally critical for inhibition of MMPs. In this region, the first 22 amino acids of the mature proteins is the most conserved among the TIMPs, 16 of the first 22 amino acids (73%) are identical among human TIMP-1, TIMP-2, and TIMP-3. However, the first 22 amino acids of mature TIMP-4 show a decreased sequence identity with other TIMPs: 63% identical to TIMP-1 and TIMP-2, and 59% identical to TIMP-3. The consensus sequence CXCXPXHPQXAFCNXDXVIRAK (single amino acid code; X = any amino acid) has been proposed to serve a diagnostic hallmark of the TIMPs being present in TIMP-1, TIMP-2, and TIMP-3 (54). Because TIMP-4 has a less conserved sequence in this region with only 12 of 22 amino acids identical in all four TIMPs, we suggest the use of consensus sequence VIRAK (positions 47-51, Fig. 2) as a diagnostic hallmark of the TIMP family. We have shown that TIMP-4 is more homologous to TIMP-2 and TIMP-3 than to TIMP-1.

Tissue expression of TIMP-4 appears to be limited. Although large amounts of transcript were detected in heart, much lower levels of expression were detected in kidney, pancreas, colon and testes; no TIMP-4 transcript were detected in other tissues such as liver, brain, lung, thymus, small intestine and spleen. TIMP-4 may function in a tissue-specific fashion as part of an acute response to tissue remodeling. It is interesting to note that the highest levels of TIMP-4 expression is seen in the heart, in which human cancer metastasis rarely occurs. The possibility that the high expression of TIMP-4 in heart may contribute the inability of malignant cells to invade needs further

consideration.

We have expressed TIMP-4 in MDA-MB-435 human breast cancer cells in an effort to investigate the biological significance of this new TIMP in tumor growth and metastasis. Since TIMPs block the activities of MMPs, the net inhibitory activity of TIMPs might be important in preventing malignant progression from the benign to the metastatic phenotype. In fact, tumor cell invasion and metastasis can be blocked by up-regulation of TIMP expression or an exogenous supply of TIMPs [17, 40-44]. Alternatively, down-regulation of TIMP-1 and TIMP-2 have been reported to contribute significantly to the invasive potential of human glioblastoma [19]. We have analyzed the MMP inhibitory activities of the expressed rTIMP-4 from the conditioned medium of transfected clones. As expected, rTIMP-4 proteins expressed from human breast cancer cells possess an inhibitory activity against MMP and are secreted extracellularly, thus confirming that the novel protein is the new member of TIMP family.

In summary, we have cloned and sequenced a novel human TIMP gene designated TIMP-4, whose expression is tissue-specific. We have also presented evidence indicating the MMP inhibitory activity of expressed TIMP-4 protein.

References

1. Dochherty, A.J.P., O'Connell, J.O., Crabbe, T., Angal, S. and Murphy, G. (1992) *Trends Biotech.* **10**, 200-207.

2. Lennarz, W.J. and Strittmatter, W.J. (1991) *Biochem Biophys. Acta* **1071**, 149-158.

3. Liotta, L.A. and Stetler-Stevenson, W.G. (1990) *Semin. Canc. Biol.* **1**, 99-106.

4. Bernhard, E.J, Gruber, S.B. & Muschel, R.J. (1994) *Proc. Nat. Acad. Sci.* **91**, 4293-4297.

5. Pyke, C., Ralfkiaer, E., Tryggvason, K. & Dano, K. (1993) *Am. J. Pathol.* **142**, 359-365.

6. Naylor, M.S., Stamp, G.W., Davies, B.D. & Balkwill, E.R. (1994) *Int. J. Cancer* **58**, 50-56.

7. Basset, P., Wolf, C., Rouyer, N., Bellocq, J.P., Rio, M.C. & Chambon, P. (1994) *Cancer* **74**, 1045-1049.

8. Polette, M., Clavel, C., Cockett, M., Girod-de-Bentzmann, S., Murphy, G. & Birembaut, P. (1993) *Invasion Metastasis* **13,** 31-37.

9. Polette, M., Gilbert, N., Stas, I., Nawrocki, B., Noel, A., Remacle, A., Stetler-Stevenson, W.G., Birembaut, P. & Foidart, M. (1994) *Virchows Arch.* **424**, 641-645.

10. Polette, M., Clavel, C., Birembaut, P. & De-Clerck, Y.A. (1993) *Pathol. Res. Pract.* **189**, 1052-1057.

11. Kossakowska, A.E., Urbanski, S.J. & and Edwards, D.R. (1991) *Blood* 77, 2475-2481.

12. Visscher, D.W., Hoyhtya, M., Ottosen, S.K., Liang, C.M., Sarkar, F.H., Crissman, J.D. & Fridman, R. (1994) *Int. J. Cancer* 59, 339-344.

13. Basset, P., Bellocq, J.P., Wolf, C., Stoll, I., Hutin, P., Limacher, J.M., Podhajcer, O.L., Chenard, M.P., Rio, M.C. & Chambon, P. (1990) *Nature* **348**, 699-704.

14. Wolf, C., Rouyer, N., Lutz, Y., Adida, C., Loriot, M., Bellocq, J.P., Chambon, P. & Basset, P. (1993) *Proc. Natl. Acad, Sci.* **90**, 1843-1847.

15. Hoyhtya, M., Fridman, R., Komarek, D., Porter-Jordan, K., Stetler-Stevenson, W.G., Liotta, L.A. & Liang, C.M. (1994) et al., *Int. J. Cancer* **56,** 500-505.

16. Poulsom, R., Hanby, A.M., Pignatelli, M., Jeffery, R.E., Longcroft, J.M., Rogers, L. & Stamp, G.W. (1993) *J. Clin. Pathol.* **46**, 429-436.

17. Khokha, R., Waterhouse, P., Yagel, S., Lala, P.K., Overall, C.M., Norton, G. & Denhardt, D.T. (1989) *Science* **243**, 947-950.

18. Denhardt, D.T., Khokha, R., Yagel, S., Overall, C.M. & Parhar, R.S. (1992) *Matrix Suppl.* **1**, 281-285.

19. Mohanam, S., Wang, S.W., Rayford, A., Yamamoto, M., Sawaya, R., Nakajima, M., Liotta, L.A., Nicolson, G.L, Stetler-Stevenson, W.G. & Rao, J.S. (1995) *Clin. Exp. Metastasis* **13**, 57-62.

20. Carmichael, D.F., Sommer, A., Thompson, R.C., Anderson, D.C., Smith, C.G., Welgus, H.G. & Stricklin, G.P. (1986) *Proc. Natl. Acad. Sci.* **83**, 2407-2411.

21. Stetler-Stevenson, W.G., Brown, P.D., Onisto, M., Levy, A.T. & Liotta, L.A. (1990) *J. Biol. Chem.* **265**, 13933-13938.

22. Wick, M., Burger, C., Brusselbach, S., Lucibello, F.C. & Muller, R. (1994) *J. Biol. Chem.* **269**, 18953-18960.

23. Uria, J.A., Ferrando, A.A., Velasco, G., Freije, J.M. & Lopez-Otin, C. (1994) *Cancer Res.* **54**, 2091-2094.

24. Wilde, C.G., Hawkins, P.R., Coleman, R.T., Levine, W.B., Delegeane, A.M., Okamoto, P.M, Ito, L.Y., Scott, R.W. and Seilhamer, J.J. (1994) *DNA and Cell Biology* **13**, 711-718.

25. Miyazaki, K., Funahashi, K., Numata, Y., Koshikawa, N., Akaogi, K., Kikkawa, Y., Yasumitsu, H. and Umeda, M. (1993) *J. Biol.Chem.* **268**, 14387-14393.

26. Apodaca, G., Rutka, J.T., Bouhana, K., Berens, M.E., Giblin, J.R., Rosenblum, M.L., McKerrow, J.H. and Banda, M.J. (1990) *Cancer Res.* **50**, 2322-2329.

27. Goldberg, G.I., Marmer, B.L., Grant, G.A., Eisen, A.Z., Wilhelm, S. and He, C.S. (189) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8207-8211.

28. Kleiner, D.E., Unsworth, E.J., Krutzsch, H.C. and Stetler-Stevenson, W.G. (1992) *Biochemistry* **31**, 1665-1672.

29. Murphy, G., Willenbrock, F., Ward, R.V., Cockett, M.I., Eaton, D. and Dochery, A.J.P. (1992) *Biochem. J.* **283**, 637-641.

30. Leco, K.J., Khokha, R., Pavloff, N., Hawkes, S.P. & Edwards, D.R. (1994) *J. Biol. Chem* **269**, 9352-9360.

31. Lotz, M. and Guerne, P.A. (1991) *J. Biol. Chem.* **266**, 2017-2120.

32. Mann, J.S., Kindy, M.S., Edwards, D.R. and Curry, T.E. (1991) *Endocrinology* **128**, 1825-1832.

33. Sato, T., Ito, A., Mori, Y., Yamashita, K., Hayakawa, T. and Nagase, H. (1991) *Biochem. J.* **275**, 645-650.

34. Edwards, D.R., Rocheleau, H., Sharma, R., Wills, A.J., Cowie, A., Hassell, J.A. and Heath, J.K. (1992) *Biochem. Biophys. Acta* **1171**, 41-55.

35. Wick, M., Burger, C., Brusselbach, S., Lucibello, F.C. and Muller, R. (1994) *J. Biol. Chem.* **269**, 18953-18960.

36. Stetler-Stevenson, W.G., Brown, P.D., Onisto, M., Levy, A.T. and Liotta, L.A. (1990) *J. Biol. Chem.* **265**, 13933-13938.

37. Denhardt, D.T., Feng, B., Edwards, D.R., Cocuzzi, E.T. and Malyankar, U.M. (1993) *Pharmacol. Ther.* **59**, 329-341.

38. Adams, M.A., Kelley, J.M., Gocayne, J.D., Dubnick, M., polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) *Science* **252**, 1651-1655.

39. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) *J. Mol. Biol.* **215**, 403-410.

40. Khokha, *R. (1994) JNCI.* **86,** 299-304.

41. Testa, J.E. (1992) *Cancer Res.* **53**, 5597-5603.

42. Alvarez, O.A., Carmichael, D.F. & DeClerck, Y.A. (1990) *JNCI* **82**, 589-595.

43. Tsuchiya, Y., Sato, H., Endo, Y., Okada, Y., Mai, M., Sasaki, T. & Seiki, M. (1993) *Cancer Res.* **53**, 1397-1402.

44. De Clerck, Y.A., Perez, N., Shimada, H., Boone, T.C., Langley, K.E. & Taylor, S.M. (1992) *Cancer Res.* **52**, 701-708.

45. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156-159.

46. Mcleskey, S.W., Kurebayashi, J., Honing, S.F., Zweibel, J.A., Lippman, M.E., Dickson, R.B. and Kern, F.G. (1993) *Cancer Res.* **53**, 2168-2177.

47. Staskus, P.W., Masiarz, F.R., Pallanck, L.J. and Hawkes, S.P. (1991) *J. Biol. Chem.* **266**, 449-454.

48. Azzam, H.S., Arand, G., Lippman, M.E. and Thompson, E.W. (1993) *JNCI* **85**, 1758-1763.

49. Williamson, R.A., Marston, F.A., Angal, S., Koklitis, P., Panico, M., Morris, H.R., Carne, A.F., Smith, B.J., Harris, T.J. & Freedman, R.B. (1990) *Biochem. J.* **268,** 267-274.

50. Woessner, J.F. Jr. (1991) *FASEB. J.* **5**, 2145-2154.

51. Murphy, G., Houbrechts, A., Cockett, M.I., Williamson, R.A., O'Shea, M. and Docherty, A.J. (1991) *Biochemistry* **30**, 8097-8102.

52. Willenbrock, F., Crabbe, T., Slocombe, P.M., Sutton, C.W., Docherty, A.J.P., Cockett, M.I., O'Shea, M., Brocklehurst, K., Phillips, I.R. and Murphy, G. (1993) *Biochemistry* **32**, 4330-4337.

53. DeClerck, Y.A., Darville, M.I., Yean, T.D., Lee, Y., Tomich, J.M. and Langley, K.E. (1993) *Biochem., J.* **289**, 65-69.

54. Apte, S.S., Olsen, B.R. and Murpthy G. (1995) *J. Biol. Chem.* **270**, 14313-14318.

55. Miyazaki, K., Funahashi, K., Numata, Y., Koshikawa, N., Akaogi, K., Kikkawa, Y., Yasumitsu, H. and Umeda, M. (1993) *J. Biol. Chem.* **268**, 14387-14393.

Figure Legends

Fig. 1. TIMP-4 cDNA sequence. The full-length cDNA was sequenced using ABI 373a Automated Fluorescent Sequencer method. The deduced amino acid sequence is shown under the DNA sequence. The translation termination codon (TAG) is labelled with *. The putative mature protein cleavage site is underlined at position 29 for alanine. Numbers refer to nucleic acid positions. The sequence has been deposited in the GenBank with an accession number of (not yet known).

Fig. 2. Comparison of the predicted amino acid sequence of human TIMP-4 with human TIMP-1, TIMP-2, and TIMP-3. The available amino acid sequence of TIMP-1 (accession number P01033), TIMP-2 (accession number P16035), and TIMP-3 (accession number P35625) were obtained from the SwissProt data base and aligned with the TIMP-4 deduced sequence using the clustal method of the MegAling Program from the DNASTAR software package. Conserved bases are boxed; the 29 amino acid putative signal sequence is shown between two triangles (▲), and the 12 conserved cysteine residues are labeled with arrows.

Fig. 3. The expression of TIMP4 gene in a variety of normal adult human tissues. Twenty micrograms of total RNA were analyzed in Northern blotting. A strong hybridizing band of 1.4 kilobase was recognized in the lane corresponding to RNA from adult heart. Additional bands with much lower intensities corresponding to mRNA species of about 4.1 kb, 2.1 kb, 1.2 kb, and 0.97 kb were also detected. The integrity of the RNA samples was ascertained by direct visualization of the ribosomal RNAs in the stained gel.

Fig. 4. Northern analysis of TIMP-4 expression in human breast cancer cells. RNAs were isolated and subjected to Northern analysis by hybridization with either a full-length cDNA probe (**A**) or a 390 base riboprobe, which represents a specific nucleotide sequence of the 3'-untranslated TIMP-4 (**C**). The integrity of the RNA was ascertained by hybridization with a house keeping gene 36B4 (**B**). Each lane contained 20 ug total RNA. Northern analysis failed to detect the TIMP-4 transcript in most breast cancer cell lines, except MDA-MB-231 cells, which showed a strong 1.4 kb TIMP-4 transcript; a very weak hybridization signal was also detected in MDA-MB-436 cells.

Fig. 5. Metalloprotease inhibitory activities produced by transforming human breast cancer cells. Human breast cancer cell line, MDA-MB-435, was transfected with either the pCITIMP4 plasmid containing the full-length TIMP-4 cDNA or the control pCI-neo plasmid, and the TIMP-4 positive clones were selected as described in "Materials and Methods". **A**. Northern blot of RNAs from both control and TIMP-4 transfected clones. Total RNAs were isolated from three control pCI-neo transfected clones (N1-N3) and four TIMP-4 transfected clones (P1-P4), and then subjected to Northern blot analysis with a random-labelled full-length TIMP-4 probe. Strong TIMP-4 transcripts were detected in 3 of 4 transfected clones; clone P3 shows low level TIMP-4 expression. In contrast, no endogenous TIMP-4 transcripts were detected in any of the control clones. The integrity of the RNAs and loading control were ascertained by hybridization with a house keeping gene 36B4 (**B**). **C**. Analysis of MMP inhibitory activity by reverse zymography. Conditioned media were prepared from one control clone N1 and two TIMP-4-producing clones P1 and P4, concentrated, and analyzed by protease-substrate gel electrophoresis as described under "Materials and Methods". Lane 1: clone P1; lane 2: clone N1; lane 3: clone P4. Arrow indicates the molecular weight of expressed TIMP-4 protein.

Acknowledgments

We thank Dr. W. G. Stetler-Stevenson for providing rMMP-2, Drs. Amy Sang and Robert Bienkowski for discussion on the manuscript and Mrs. Jane Shirreffs for reading the manuscript. This work was supported by grant CA68064-01 from National Institute of Health, grant DAMD17-94-J-4149 from U.S. Department of Army, and the Irving and Helen Schneider family.

Fig. 1. TIMP-4 cDNA sequence. The full-length cDNA was sequenced using ABI 373a Automated Fluorescent Sequencer method. The deduced amino acid sequence is shown under the DNA sequence. The translation termination codon (TAG) is labelled with *. The putative mature protein cleavage site is underlined at position 29 for alanine. Numbers refer to nucleic acid positions. The sequence has been deposited in the GenBank with an accession number of (not yet known).

Figure 1 not available at time manuscript was prepared

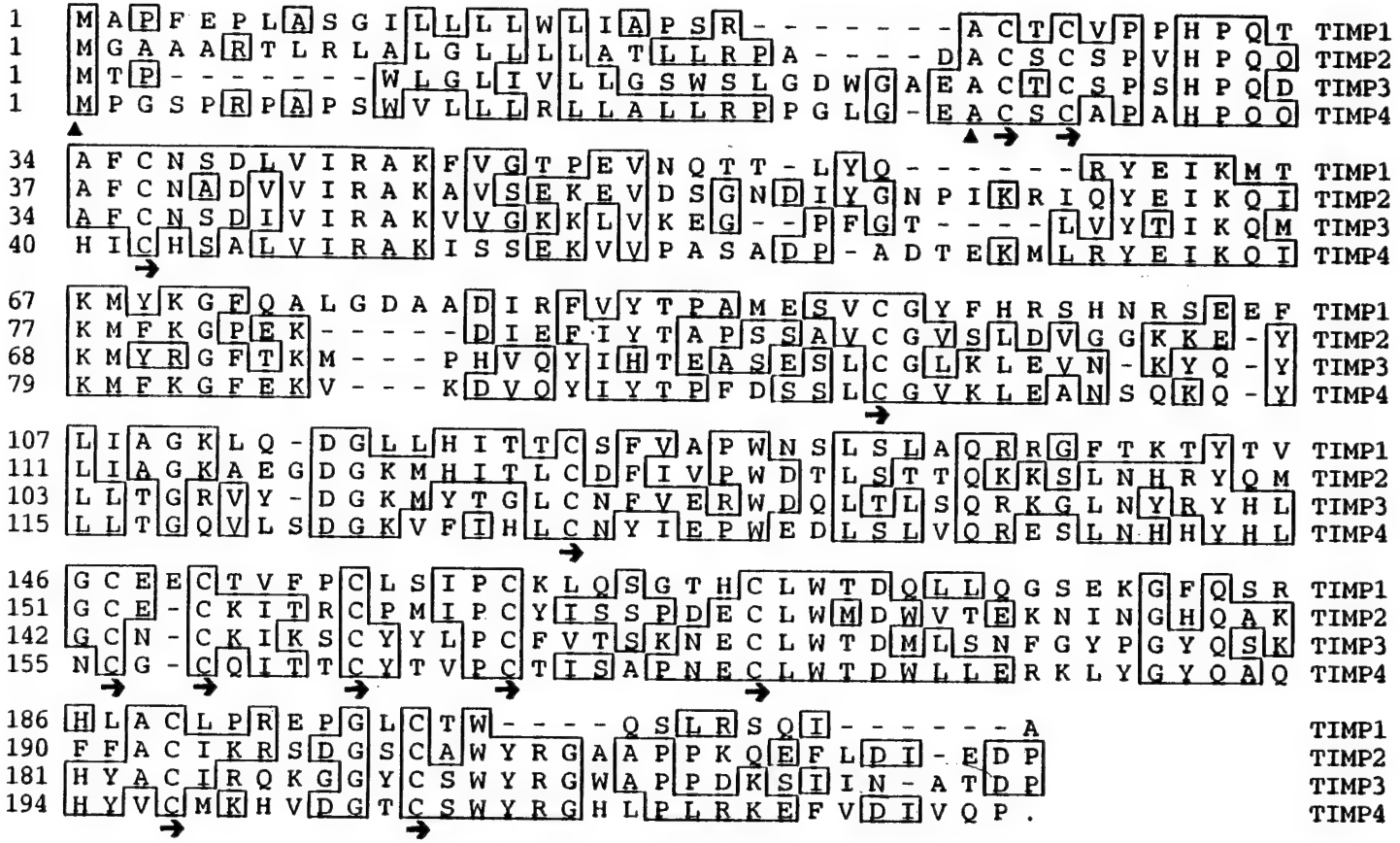

1 MAPFEPLASGILLLLWLIAPSR-------ACTCVPPHPQT TIMP1
1 MGAAARTLRLALGLLLLLATLLRPA----DACSCSPVHPQQ TIMP2
1 MTP--------WLGLIVLLGSWSLGDWGAEACTCSPSHPQD TIMP3
1 MPGSPRPAPSWVLLLRLLALLRPPGLG-EACSCAPAHPQQ TIMP4
34 AFCNSDLVIRAKFVGTPEVNQTT-LYQ-------RYEIKMT TIMP1
37 AFCNADVVIRAKAVSEKEVDSGNDIYGNPIKRIQYEIKQI TIMP2
34 AFCNSDIVIRAKVVGKKLVKEG--PFGT----LVYTIKQM TIMP3
40 HICHSALVIRAKISSEKVVPASADP-ADTEKMLRYEIKQI TIMP4
67 KMYKGFQALGDAADIRFVYTPAMESVCGYFHRSHNRSEEF TIMP1
77 KMFKGPEK-----DIEFIYTAPSSAVCGVSLDVGGKKE-Y TIMP2
68 KMYRGFTKM---PHVQYIHTEASESLCGLKLEVN-KYQ-Y TIMP3
79 KMFKGFEKV---KDVQYIYTPFDSSLCGVKLEAN-SQKQ-Y TIMP4
107 LIAGKLQ-DGLLHITTCSFVAPWNSLSLAQRRGFTKTYTV TIMP1
111 LIAGKAEGDGKMHITLCDFIVPWDTLSTTQKKSLNHRYQM TIMP2
103 LLTGRVY-DGKMYTGLCNFVERWDQLTLSQRKGLNYRYHL TIMP3
115 LLTGQVLSDGKVFIHLCNYIEPWEDLSLVQRESLNHHYHL TIMP4
146 GCEECTVFPCLSIPCKLQSGTHCLWTDQLLQGSEKGFQSR TIMP1
151 GCE-CKITRCPMIPCYISSPDECLWMDWVTEKNINGHQAK TIMP2
142 GCN-CKIKSCYYLPCFVTSKNECLWTDMLSNFGYPGYQSK TIMP3
155 NCG-CQITTCYTVPCTISAPNECLWTDWLLERKLYGYQAQ TIMP4
186 HLACLPREPGLCTW----QSLRSQI-------A TIMP1
190 FFACIKRSDGSCAWYRGAAPPKQEFLDI-EDP TIMP2
181 HYACIRQKGGYCSWYRGWAPPDKSII N-ATDP TIMP3
194 HYVCMKHVDGTCSWYRGHLPLRKEFVDIVQP. TIMP4

Figure 2

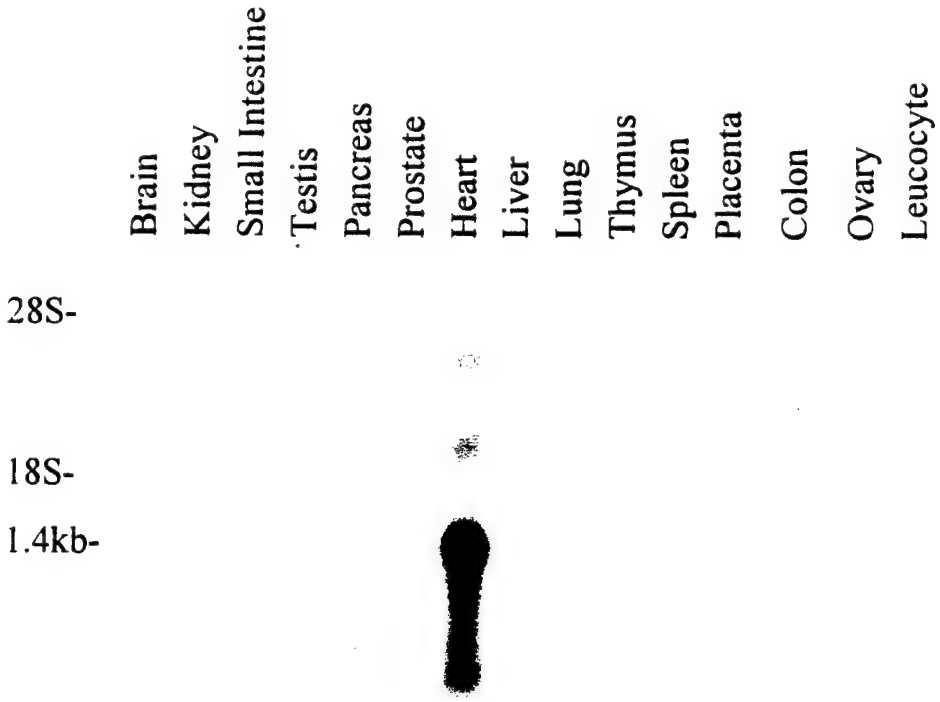


Figure 3

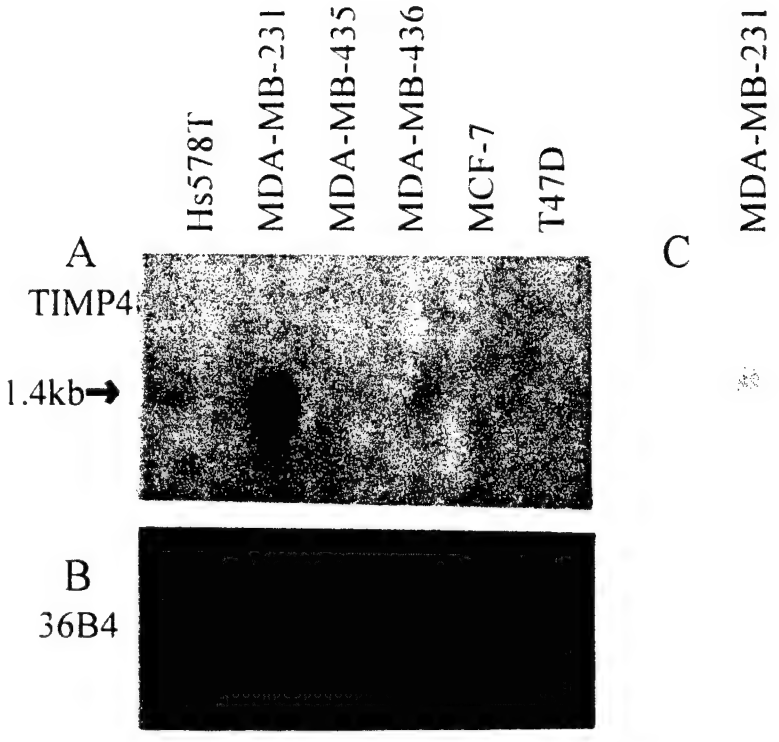
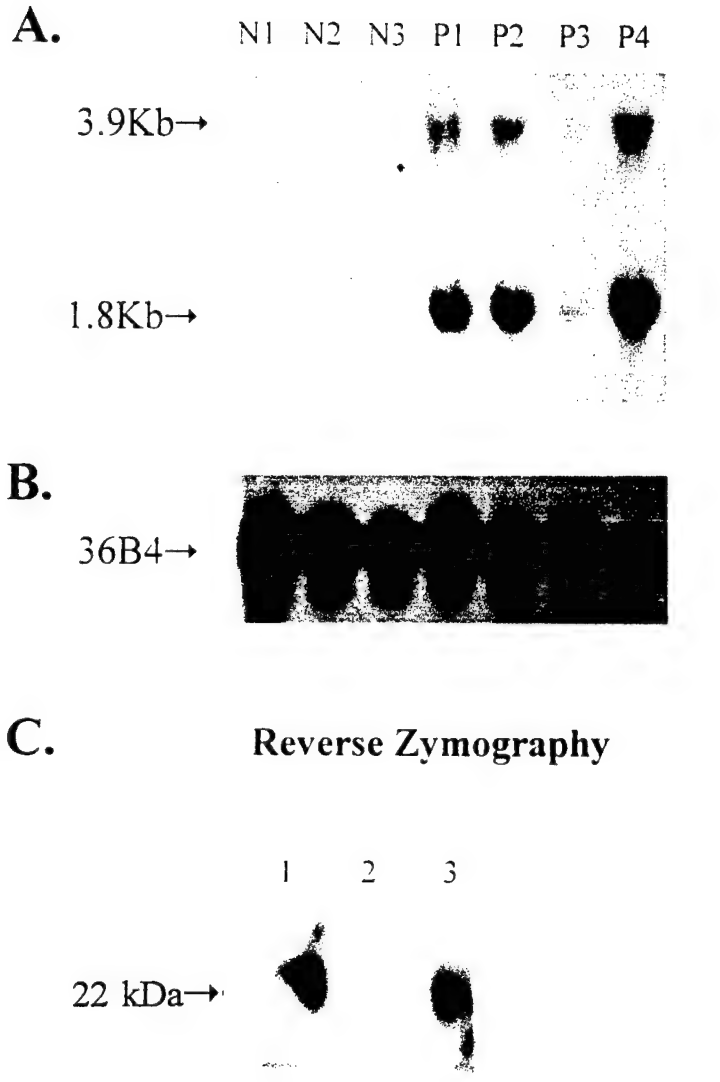


Figure 4

Figure 5


A.
N1 N2 N3 P1 P2 P3 P4
3.9Kb→
1.8Kb→
B.
36B4→
C.
Reverse Zymography
1 2 3
22 kDa→

Identification of a Breast Cancer Specific Gene, BCSG1, by Direct Differential cDNA Sequencing

Yuenian Eric Shi[1,4], Mingsheng Wang[1], Tongli Jia[1], Jingwen Liu[3], Alexander Fuchs[4], Craig Rosen[2]

Yiliang E. Liu[1] and [1]Hongjun Ji[2]

Departments of [1]Pediatrics and [4]Pathology

Long Island Jewish Medical Center,

The Long Island Campus for Albert Einstein College of Medicine

New Hyde Park, NY 11042

[2]Human Genome Sciences, Inc., Rockville, MD 20850-3338.

[3]Mountain States Medical Research Institute and Department of Veterans Affairs Medical Center

Boise, Idaho 83702

Corresponding author

Dr. Yuenian Eric Shi

Pediatric Research Center, Schneider Children Hospital, Long Island Jewish Medical Center

New Hyde Park, NY 11042

Phone 718-470-3086; FAX 718-470-6744; e-mail Shi@aecom.yu.edu

Running title: A novel putative breast cancer specific gene, BCSG1

Abstract

A high-throughput direct differential cDNA sequencing approach was employed to identify genes differentially expressed in normal breast as compared with breast cancer. Approximately six thousand expressed sequence tags (EST) from complementary DNA (cDNA) libraries of normal breast and breast carcinoma were randomly selected and subjected to EST sequencing analysis. The relative expression levels of more than 2,000 unique EST groups were quantitatively compared in normal versus cancerous breast. Of many putative differentially expressed genes, a breast cancer specific gene BCSG1, which was expressed in high abundance in a breast cancer cDNA library but scarcely in a normal breast cDNA library, was identified as a putative breast cancer marker. *In situ* hybridization analysis demonstrated a stage-specific BCSG1 expression: undetectable in normal or benign breast lesions, partial expression in ductal carcinoma *in situ*, but extremely high level in advanced infiltrating breast cancer. The predicted amino acid sequence of BCSG1 gene has a significant sequence homology to non-Aß fragment of Alzheimer's disease amyloid protein. BCSG1 overexpression may indicate breast cancer malignant progression from benign breast or *in situ* carcinoma to the highly infiltrating carcinoma.

INTRODUCTION

The onset and progression of breast cancer is accompanied by multiple genetic changes that result in qualitative and quantitative alterations in individual gene expression (1). Our hypothesis is that many of these quantitative genetic changes manifest themselves as alterations in the cellular complement of novel transcribed mRNAs. Identification of these mRNAs, if sufficiently characterized, could provide clinically useful information for patient management and prognosis while enhancing our understanding of breast cancer pathogenesis. Although pathological endpoints such as tumor size, lymph node status and status of estrogen receptor and progesterone receptor remain the most useful guides in prognosis and selecting treatment strategies for breast cancer (2), there is a need to further investigate the molecular mechanisms that determine the properties of an individual tumor e.g., probability of metastasis. While numerous prognostic factors have now been identified, few have contributed to defining clinical response to therapy.

Identification of quantitative changes in gene expression that occur in the malignant mammary gland, if sufficiently characterized, may yield novel molecular markers which may be useful in the diagnosis and treatment of human breast cancer. Several differential cloning methods, such as differential display polymerase chain reaction and subtractive hybridization, have been used to identify the genes differentially expressed in breast cancer biopsies, as compared to normal breast tissue controls (3-7). However, these investigations have involved the relatively time- and labor-intensive steps of subcloning, library screening, and cDNA sequencing of individual genes (4,8). On the other hand, creation of expressed sequence tag libraries is a rapid method used to identify or "tag" sequences that are expressed in specific tissues (9-10). Since the introduction of the EST sequencing approach, many novel human genes have been discovered (9-10). The advantage of this methodology, compared to isolation and sequencing of individual cDNAs, is that a large number of sequences can be

"cataloged" with small amounts of sequencing data.

With the availability of tens of thousands of ESTs, researchers now shift their attention to the unveiling of expression profile of individual genes or pattens of genes in normal versus diseased states. Several newly developed strategies, such as Serial Analysis of Gene Expression (SAGE) (11) and cDNA Microarray method (12), have demonstrated potential for broad application for quantitative analysis of differential patterns of gene expression. Within this context, we undertook a search, using the differential cDNA sequencing approach, for isolation of differentially expressed sequence tags and the possible presence of the new marker genes for breast cancer. In this initial report, we described a novel breast cancer specific gene named BCSG1 that is overexpressed in advanced infiltrating breast cancer cells but not in normal or benign breast lesion. The expression pattern of BCSG1 may be a meaningful marker in the development of breast cancer.

MATERIALS AND METHODS

Reagents. Restriction enzymes, T7 polymerase, random primer DNA labeling kit, and digoxigenin-labeled nucleotides were obtained from Boehringer Mannhem, Indianapolis. ^{32}P-dATP was purchased from Amersham.

Differential cDNA Sequencing. We have used EST analysis to search for new genes differentially expressed in breast cancer versus normal breast. A data base containing approximately 500,000 human partial cDNA sequences (expressed sequence tags) has been established in a collaborative effort between the Institute for Genomic Research and Human Genome Science Inc., using high throughput automated DNA sequence analysis of randomly selected human cDNA clones (10). RNAs from a stage III breast carcinoma and patient-matched normal breast were isolated and subjected to preparation of cDNA libraries. EST automated DNA sequence analysis was performed on randomly selected cDNA clones. Both libraries had about 60% novel gene sequences which did not match exactly to published human genes. A total of 3048 ESTs from breast cancer cDNA library and 2886 ESTs from normal breast cDNA library were randomly picked and sequence analyzed. The ESTs with overlapping sequences were grouped into unique EST groups; and each EST group may represent a gene or a family of sequence-related genes. There were more than 2,200 EST groups that were analyzed for quantitative comparison of EST hits in the pair of cDNA libraries from normal breast versus breast cancer by examining the expression of individual EST sequences. The numbers of EST hits in the libraries reflect the relative expression or mRNA transcript copy numbers of the EST. This direct differential cDNA sequence, as illustrated in Fig. 1, utilizing the direct EST sequencing analysis simultaneously on a pair of cDNA libraries made from normal breast and breast cancer, was used to study expression profile of individual genes and patterns of genes in normal breast versus breast cancer.

Tissue-Specific Expression Analysis. Analysis of relative expression of breast-derived ESTs versus their expression in other tissues was performed. The differentially expressed EST groups identified by differential cDNA sequence were analyzed for tissue-specific expression against the total of 500,000 ESTs from a variety of different cDNA libraries.

Northern analysis. Total RNA was extracted from tissues according to the method of Chomcznski and Sacchi (45). The RNA from human breast cancer cells was prepared using the RNA isolation kit RNAzol B (Tel-Test, Inc) based on the manufacturer's instruction. Equal aliquots of RNA were electrophoresed in a 1.2% agarose gel containing formaldehyde and transferred to nylon membrane (Boehringer Mannheim). The membrane was pre-hybridized with ExpressHyb hybridization solution (Clontech, Inc.) at 68^{0}C for 30 min. The hybridization was carried out in the same solution with ^{32}P-labeled BCSG1 probe (1.5 x 10^6 cpm/ml) for 1 hour at 68^{0}C. The membrane was then rinsed in 2 x SSC containing 0.05% SDS three times for 30 min at room temperature, followed by two washes with 0.1 x SSC containing 0.1% SDS for 40 min at 50^{0}C. The full-length BCSG1 cDNA was isolated from the Bluescript vector, following EcoRI and XhoI digestion, and used as a template for preparation of a random-labelled cDNA probe.

***In situ* hybridization.** *In situ* hybridization was carried out as described (13). Briefly, deparaffinized and acid-treated sections (5-um thick) were treated with proteinase K, pre-hybridized, and hybridized overnight with digoxigenin labeled anti-sense transcripts from a BCSG1 cDNA insert. The BCSG1 antisense probe is a 550 bp full-length fragment. The probe was generated by PstI cut of BCSG1 cDNA plasmid and followed by T7 polymerase. Hybridization was followed by RNase treatment and three stringent washings. Sections were incubated with mouse anti-digoxigenin antibodies (Boehringer) followed by the incubation with biotin-conjugated secondary rabbit anti-mouse antibodies (DAKO). The

colorimetric detection were performed using a standard indirect streptavidin-biotin immunoreaction method by DAKO's Universal LSAB Kit according to manufacturer's instructions.

RESULTS

Molecular cloning of BCSG1 complementary DNA. We generated cDNA libraries from breast cancer biopsy specimen and patient-matched normal breast and analyzed these libraries by EST sequencing. Approximately 6,000 ESTs were analyzed and grouped to different groups based on sequence overlapping, and 2,200 unique EST groups were first analyzed for relative expression in the cDNA libraries from normal breast versus breast cancer and then subjected to tissue-specific expression by examining tissue origins of individual EST sequences against a large population of ESTs derived from a variety of different tissue types. As a result, we identified three classes of EST groups that were differentially expressed in normal breast versus breast cancer. As a demonstration of this approach, Table 1 shows a partial list of three classes of genes that are differentially expressed in normal breast versus breast cancer. Class I represents the genes more abundant in breast cancer than in normal breast and includes cathepsin D, a well-studied steroid regulated extracellular matrix-degrading proteinase (14-16). Cathepsin D is thought to play a role in breast cancer metastasis (14-16) and has been proposed as a prognostic marker in breast cancer progression (17-19,38). As listed, there were 5 cathepsin D ESTs sequenced in the breast cancer cDNA library and only 1 EST in the normal breast cDNA library. Another proposed breast cancer metastasis-related gene and a prognostic marker for breast cancer, 67 kDa laminin receptor (20-24), was also picked up in this class by the differential cDNA sequencing approach. Class II represents genes that are more abundant in normal breast than in breast cancer.

Although the genes in classes I and II are differentially expressed in normal breast versus breast cancer, these genes are unique to breast tissues. Class III is a special group of genes that are selectively expressed in breast relative to other tissue types. The tissue-specific expression of the unique gene was searched against approximately 500,000 ESTs using the BLAST program (25). None of these breast cancer specific genes (BCSG) except the first one matched with any sequences in public gene sequence

databases. BCSG1 was chosen for analysis as a first putative breast cancer maker gene because 1) its sequence has been matched with the sequence in public gene sequence database; and 2) most of the individual EST sequences in BCSG1 were derived from a breast tumor cDNA library. Of the eight distinctive EST clones in BCSG1, seven of them were discovered in breast cDNA libraries and only one in a brain library. Of the seven EST clones discovered in the breast cDNA libraries, six of them were identified in the breast tumor library and only one in the normal breast library. After sequencing analysis of all 6 EST clones, one EST clone was found to have a complete full-length sequence. The open reading frame of the resulting full-length gene is predicted to encode a 127 amino acid polypeptide. Comparison of the predicted amino acid sequence with the sequence of a similar human protein is shown in Fig. 2. After optimal alignment, the putative BCSG1-encoded protein shows 54% sequence identity with the recently cloned non-Aß fragment of human Alzheimer's disease (AD) amyloid protein (26).

Tissue expression. The expression of BCSG1 gene in a variety of normal human tissues were analyzed by Northern blotting (Fig. 3). As expected, the Northern blot showed that BCSG1 was abundantly expressed as a 1 kb transcript in brain which is the rich source for AD amyloid family gene. Similar bands with much lower accumulations in their relative intensity were also obtained in ovary, testis, colon, and heart. By contrast, none of them was present in other specimens analyzed, such as breast, kidney, liver, prostate, lung, small intestine, thymus and placenta.

Expression of BCSG1 in human breast cancer cells. In an attempt to evaluate the potential biological significance of BCSG1 on human breast cancer development and progression, we studied BCSG1 gene expression in human breast biopsy samples. The expression of BCSG1 in metastatic breast carcinoma and normal breast tissue were analyzed by Northern blotting. Fig. 4 shows overexpression

of BCSG1 transcript in an infiltrating breast carcinoma. In contrast, the BCSG1 transcript was undetectable in normal breast tissue. The presence of BCSG1 transcript in human breast tissue and its overexpression in breast carcinomas are consistent with our differential cDNA sequencing cloning strategy which suggests a possible role or a biomarker of up-regulation of BCSG1 in the development of breast cancer.

The expression of BCSG1 was also investigated in a variety of human breast cancer cell lines (Fig. 5). Northern blot detected the 1 Kb BCSG1 transcript in 2/4 lines derived from pleural effusion and 4/4 lines detected from ductal infiltrating carcinomas. Among these lines, H3922 expressed the highest level of BCSG1 mRNA. The absence of BCSG1 mRNA in some breast cancer cell lines may suggest that the expression of BCSG1 gene requires specific in vivo conditions or that it is induced by interactions between the tumor cells and stromal cells.

In order to localize the cellular source of the BCSG1 expression and to further assess the biological relevance of the overexpression of BCSG1 in breast cancers, we next performed *in situ* hybridization on fixed breast sections from 20 infiltrating carcinomas, 15 ductal carcinomas *in situ* (DCIS), and 18 benign breast lesions including 5 reduction mammoplasty specimens, 8 breast hyperplasias, and 5 fibroadenomas. In these experiments, we examined two aspects of BCSG1 expression: 1) the tissue localization (stromal versus epithelial); and 2) the correlation of BCSG1 expression and breast cancer malignant phenotype. There was a wide variation in staining intensity for BCSG1 expression among the human breast cancer specimens. Since the colorimetric *in situ* hybridization is not quantitative, the tissue samples were classified into either positive or negative staining for BCSG1 expression; no attempt was made to differentiate the levels of expression of BCSG1 among positive-staining specimens. The negative cases were confirmed with at least two independent experiments. All stainings were reviewed by at least two people. Fig. 6 shows a representative *in situ*

hybridization for BCSG1. We found a strongly positive BCSG1 hybridization in neoplastic epithelial cells of highly infiltrating breast carcinomas (Fig. 6A,B). The expression of BCSG1 mRNA was detectable in the neoplastic epithelial cells in 17 of 20 infiltrating breast carcinomas. No expression of BCSG1 was detected in the stromal cells. In contrast, expression of BCSG1 was absent in 16 out of 18 cases of normal or benign breast lesions. A representative negative staining of BCSG1 in normal ductal breast epithelial cells (Fig. 6E), a benign proliferative breast lesion (Fig. 6F), and a benign fibroadenoma (Fig. 6G) are presented. Furthermore, as demonstrated in Fig. 6B for a highly invasive breast carcinoma, no detectable signal of BCSG1 expression was evident in the residual normal lobular breast epithelial cells although the surrounding invasive breast carcinoma cells were stained positive for BCSG1 expression. These *in situ* hybridization results are consistent with the Northern blot analysis which showed a strong expression of BCSG1 transcript in breast carcinoma but not in normal or benign breast lesions.

It is interesting to note that although a strong BCSG1 signal was easily detected in the malignant breast epithelial cells of infiltrating breast carcinoma, the *in situ* carcinomas showed a different BCSG1 expression patterns. Among 15 DCISs (8 are Comedo type and 7 are non-Comedo type), 8 specimens were stained negatively (Fig. 6D) and 7 specimens were positive (Fig. 6C). Interestingly, 6 of 7 BCSG1 positive DCIS samples were Comedo type DCIS and only one was non-Comedo type; among the BCSG1 negative specimens, there were 6 non-Comedo type DCISs and only two Comedo type DCISs. These results, which demonstrated a stage-specific BCSG1 expression from virtually no detectable expression in normal or benign breast to partial expression (7/15) in the *in situ* breast carcinoma and to the high expression (17/20) in the infiltrating malignant breast carcinomas, suggest an association of BCSG1 expression with breast cancer malignant progression. Based on this BCSG1 expression pattern, we propose that BCSG1 may be potentially used as a breast cancer progression marker.

DISCUSSION

More than 190,000 new cases of breast cancer are diagnosed in the United States every year, with incidence increasing by approximately 1% annually (27-28). Studies linked to the discovery of new genetic markers will provide new information leading to understanding of breast cancer development and progression. There are two classes of genes affecting tumor development. Genes influencing the cancer phenotype that act directly as a result of changes (eg., mutation) at the DNA level, such as BRCA1, BRCA2, and p53, are called Class I genes. The Class II genes affect the phenotype by modulation at the expression level. Development of breast cancer and subsequent malignant progression is associated with alterations of a variety of genes of both classes. Many new predictive and prognostic factors have been proposed and studied for breast cancer. HER 2/neu positive tumors respond poorly to endocrine treatment (29-30). p53 alteration has an indication of poorer prognosis and poor response to tamoxifen (31-32). The lack of Nm23 expression has an indicative value of metastatic potential and poor prognosis in invasive ductal carcinoma (33). Cathepsin D, a protease suggested to have a role in breast cancer, appears to affect the potential for invasive growth (11-13,34). Positive immunostaining of tumor sections with Factor VIII antibodies seems to be a marker for angiogenesis (35-37). It has been postulated that these tumors are targets for anti-angiogenesis drug treatment. Expression of the *mdr-1* gene is proposed to be an indicator of multidrug resistance (36-37). Poor response to endocrine therapy has been indicated for uPA/PAI-1, a plasminogen activator/inhibitor (38). Also receiving major attention are the familial breast cancer related genes, BRCA1 and BRCA2 (39-41). With the availability of tens of thousands of EST sequences, we have, using differential cDNA sequence, identified a new putative breast cancer marker gene BCSG1 and studied its expression in breast cancer.

The differential cDNA sequencing method described here is a direct approach that utilizes an

automatic EST analysis on a pair of cDNA libraries. Unlike previously described methods, the differential cDNA sequencing approach allows one to identify differentially expressed genes or pattens of genes directly from computer database. With the advancement of more efficient and rapid sequencing technology, the direct differential cDNA sequencing approach may offer a powerful method for simultaneous analysis of the expression profile of thousands of genes, as well as for the discovery of novel genes of clinical interest.

Using *in situ* hybridization analysis, we have demonstrated the expression of BCSG1 transcripts in the neoplastic epithelial cells of infiltrating breast carcinoma but not in epithelial cells of normal and benign breast. The overexpression (17 of 20) of BCSG1 in malignant infiltrating breast epithelial cells compared to the partial expression (7 of 15) in the *in situ* carcinoma suggests that up-regulation of BCSG1 expression is associated with breast cancer malignant progression and may signal the more advanced invasive/metastatic phenotype of human breast cancer. This implication is further supported by detection of BCSG1 expression in 6/8 aggressive Comedo type DCISs and only 1/7 in non-Comedo type DCISs. It is unlikely that BCSG1 is overexpressed as a secondary effect of cellular proliferation because no detectable BCSG1 expression is evident in rapidly proliferating nonmalignant breast lesions (Fig. 6F).

It will be interesting to investigate if BCSG1 expression in DCIS may indicate a malignant progression leading to invasion and metastasis. There is cause for concern about the large number of DCIS cases that are being diagnosed as a consequence of screening mammography, most of which are treated by some form of surgery. In addition, the proportion of cases treated by mastectomy may be inappropriately high (28). DCIS by definition has intact basement membrane by light microscopy (47). Defective basement membrane, however, have been found when they are stained with periodic acid-Schiff reagent and when they are examined by electron microscopy (48). In fact, it has been reported that re-evaluation by experienced pathologists showed that 28 and 15 percent of previously diagnosed

DCIS demonstrated invasion (49-50). If BCSG1 expression can provide some prognostic information on distinguishing the DCIS which is not likely to become invasive from the DCIS which is most likely to become invasive, this will help to direct the treatment strategies and to reduce some inappropriate or unnecessary mastectomies.

It is interesting to note that the predicted amino acid sequence of BCSG1 gene shares high sequence homology with non-Aß component of Alzheimer's disease (AD) amyloid precursor protein (26). A neuropathological hallmark of AD is a widespread amyloid deposition resulting from beta-amyloid precursor proteins (beta APPs). Beta AAPs are large membrane-spanning proteins that either give rise to the beta A4 peptide (Aß fragment) (42) or a non-Aß component of AD amyloid (26) that is either deposited in AD amyloid plaques or yielding soluble forms. While the insoluble membrane-bound AD amyloid destabilizes calcium homeostasis and thus renders cell vulnerable to excitotoxic conditions of calcium influx resulting from energy deprivation or overexcitation (43), the soluble AD amyloid proteins are neuroprotective against glucose deprivation and glutamate toxicity, perhaps through their ability to lower the intraneuronal calcium concentration (44). We currently do not know whether BCSG1 is an instigator or a by-product during breast cancer progression. With the availability of anti-BCSG1 antibody to localize BCSG1 protein and the recombinant BCSG1 protein, we may start to speculate that BCSG1, like soluble AD amyloid, may be potentially involved in protection of tissue damage resulting from tissue remodeling due to the local cancer invasion. An elucidation of the reasons for BCSG1 overexpression in infiltrating breast cancer cells may shed some light on the pathogenesis of breast cancer progression. Nevertheless, we demonstrated a stage-specific BCSG1 expression and an association of BCSG1 overexpression with clinical aggressiveness of breast cancers. The notion that the BCSG1 overexpression may indicate breast cancer malignant progression from benign breast or *in situ* carcinoma to the highly infiltrating carcinoma warrants further investigation.

Acknowledgments

We thank Drs. W. Cance and E. Liu of North Carolina University for providing normal breast tissue and breast cancer for constructing cDNA libraries, Dr. G. Yu and L. Xing for technical assistance, and Ms. Jane Shirreffs for reading the manuscript. The sequence analysis was conducted at the DNA Sequencing Core Facility at Human Genome Science, Inc. and The Institute of Genomic Research. Supported in part by grants from US Army Breast Cancer Research Program (DAMD17-94-J-4149) and Irving and Helen Schneider's family.

References:

1. Porter-Jordan, K. and Lippman, M.E. Overview of the biological markers of breast cancer. Hematol. Oncol. Clin. N. Am., *8*: 73-100, 1994.

2. Manning, D.L., McClelland, R.A., Knowlden, J.M., Bryant, S., Green, C.D., Sutherland, R.L. and Ormandy, C.J. Differential expression of oestrogen regulated genes in breast cancer. Acta Oncol., *34*: 641-646, 1995.

3. Watson, M.A. and Fleming, T.P. Isolation of differentially expressed sequence tags from human breast cancer. Cancer Res., *54*: 4598-4602, 1994.

4. Sager, R., Anisowicz, A., Neveu, M., Liang, P. and Sotiropoulou, G. Identification by differential display of alpha 6 integrin as a candidate tumor suppressor gene. FASEB J., *7*: 964-970, 1993.

5. Chen, Z. and Sager, R. Differential expression of human tissue factor in normal mammary epithelial cells and in carcinomas. Mol. Med., *1*: 153-160, 1995.

6. Zhang, M., Zou, Z., Maass, N. and Sager, R. Differential expression of elafin in human normal mammary epithelial cells and carcinomas is regulated at the transcriptional level. Cancer Res., *55*: 2537-2541, 1995.

7. Zou, Z., Anisowicz, A., Hendrix, M.J., Thor, A., Neveu, M., Sheng, S., Rafidi, K., Seftor, E. and Sager R. Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. Science, *263*: 526-529.

8. Liang, P., Averboukh, L., Keyomarsi, K., Sager, R. and Pardee, A.B. Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. Cancer Res., *52*: 6966-6968, 1992.

9. Adams, M.D., Soares, M.B., Kerlavage, A.R., Field, C. and Venter, J.C. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nat. Genet., *4*: 373-380, 1993.

10. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Oide, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., and Venter, J.C. Complementary DNA sequencing: expressed sequence tags and human genome project. Science, *252*: 1651-6, 1991.

11. Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. Serial analysis of gene expression. Science, *270*: 484-7, 1995.

12. Schena, M., Shalon, D.,Davis, R.W. and Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, *270*: 467-70, 1995.

13. M.L. Angerer and R.C. Angerer, In: D. Rickwood and B.D. Hames (ed.), *In situ hybridization,* pp. 15-32, London: LRL Press., 1992.

14. Rochefort, H., Capony, F., Garcia, M., Cavailles, V., Freiss, G., Chambon, M., Morisset, M. and Vignon, F. Estrogen-induced lysosomal proteases secreted by breast cancer cells: a role in carcinogenesis? J.Cell. Biochem., *35*: 17-29, 1987.

15. Cavailles, V., Augereau, P. and Rochefort, H. Cathepson D gene of human MCF-7 cells contains estrogen-responsive sequences on its 5' proximal flanking region. Biochem. Biophys. Res. Commun, *174*: 816-24, 1991.

16. Capony,F., Rougeot, C., Cavailles, V. and Rochefort, F. Estradiol increases the secretion by MCF-7 cells of several lysosomal pro-enzymes. Biochem. Biophys. Res. Commun, *171*: 972-8, 1990.

17. Brouillet, J.P., Theillet, C., Maudelonde, T., Defrenne, A., Simony-Lafontaine, J., Sertour, J., Pujol, H., Jeanteur, P. and Rochefort, H. Cathepsin D assay in primary breast cancer and lymph nodes: relationship with c-myc, c-erb-B-2 oncogene amplification and node invasiveness. Eur. J. Cancer, *26*: 437-41, 1990.

18. Spyratos, F., Maudelonde, T., Brouillet, J.P., Brunet, M., Defrenne, A., Andrieu, C., Hacene, K., Desplaces, A., Rouesse, J. and Rochefort, H. Cathepsin D: an independent prognostic factor for

metastasis of breast cancer. Lancet, *11*: 1115-8, 1989.

19. Rochefort, H., Cavailles, V., Augereau, P., Capony, F., Maudelonde, T., Touitou, I. and Garcia, M. Overexpression and hormonal regulation of pro-cathepsin D in mammary and endometrial cancer. J. steroid. Biochem.,*34*: 177-82, 1989.

20. Horan-Hand, P., Colcher, D., Salomon, D., Ridge, J., Noguchi, P. and Schlom, J. Influence of spatial configuration of carcinoma cell populations on the expression of a tumor-associated glycoprotein. Cancer Res., *45*: 833-40, 1986.

21. Hunt, G. The role of laminin in cancer invasion and metastasis. Exp. Cell Biol., *57(3)*: 165-76, 1989.

22. Castronovo, V., Colin, C., Claysmith, A.P., Chen, P.H., Lifrange, E., Lambotte, R., Krutzsch, H., Liotta, L.A. and Sobel, M.E. Immunodetection of the metastasis-associated laminin receptor in human breast cancer cells obtained by fine-needle aspiration biopsy. Am. J. Pathol., *137(6)*: 1373-81, 1990.

23. Marques, L.A., Franco, E.L., Torloni, H., Brentani, M.M., da-Silva-Neto, J.B., and Brentani, R.R. Independent prognostic value of laminin receptor expression in breast cancer survival. Cancer-Res., *50(5)*: 1479-83, 1990.

24. Gasparini, G., Barbareschi, M., Boracchi, P., Bevilacqua, P., Verderio, P., Dalla-Palma, P. and Menard, S. 67-kDa laminin-receptor expression adds prognostic information to intra-tumoral microvessel density in node-negative breast cancer. Int. J. Cancer., *60(5)*: 604-10, 1995.

25. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. Basic local alignment search tool. J. Mol. Biol., *215(3)*: 403-10, 1990.

26. Ueda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A, Yoshimoto, M., Otero, D.A., Kondo, J., Ihara, Y. and Saitoh, T. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. Proc. Natl. Acad. Sci. U.S.A., *90(23)*: 11282-6, 1993.

27. Goldhirsch, A., Commentary: Meeting highlights (St Gallen 1995): International Consensus Panel on the Treatment of Primary Breast Cancer. JNCI, *97*: 1141-1145, 1995.

28. Ernster, V.L., Barclay, J., Kerlikowske, K., Grady, D., Henderson, C. Incidence of and treatment for ductal carcinoma in situ of the breast. JAMA, *275*: 913-918, 1996.

29. Allred DC, Clark GM, Tandon AK, Molina R, Tormey DC, Osborne CK, Gilchrist KW, Mansour EG, Abeloff M, Eudey L. Her-2/neu in node-negative breast cancer. Prognostic significance of overexpression influenced by the presence of in situ carcinoma. J Clin Oncol., *10*: 599-605, 1992.

30. Gusterson BA, Gelber RD, Goldhirsch A, Price KN, Save Soderborgh J, Anbazhagan R, Styles J, Rudenstam CM, Golouh R, Reed R. Prognostic importance of c-erbB-2 expression in breast cancer. J Clin Oncol.. *10*: 1049-56, 1992.

31. Bergh J, Norberg T, Sjogren S, Lindgren A, Holmberg L. Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. Nature Medicine, *10*: 1029-34, 1995.

32. Elledge RM, Fuqua SA, Clark GM, Pujol P, Allred DC. The role and prognostic significance of p53 gene alterations in breast cancer. Breast Cancer Res Treat., *27*: 95-102, 1993.

33. Steeg PS, De La Rosa A, Flatow U, Macdonald NJ, Benedict M, Leone A. Nm23 and breast cancer metastasis. Breast Cancer Res Treat., *25*: 175-87, 1993.

34. Ferno M, Baldetorp B, Borg A, Brouillet JP, Olsson H, Rochefort H, Sellberg G, Sigurdsson H, Killander D. Cathepsin D, both a prognostic factor and a predictive factor for the effect of adjuvant tamoxifen in breast cancer, South Sweden Breast Cancer Group. Eur J Cancer, *30A*: 2042-8,1994.

35. Klijn JGM, Berns EMJJ, Foekens JA. Prognostic factors and response to therapy in breast cancer. Cancer Surveys: Breast cancer *18*: 165-98, 1993.

36. Harris AL, Fox S, Leek R, Zhang H, Scott P, Bicknell R, Gatter K. Breast cancer angiogenesis. Therapy target and prognostic factor. Eur J Cancer, *31A*: 831-2, 1995.

37. Gasparini G, Pozza F, Hariss AL. Evaluating the potential usefulness of new prognostic and predictive indicators in node-negative breast cancer patients. JNCI, *85*: 1206-19, 1993 (errata JNCI 1993; 85:1605).

38. Foekens JA, Look MP, Peters HA, Van Putten WL, Portengen H, Klijn JG. Urokinase-type plasminogen activator and its inhibitor PAI-1: predictor of poor response to tamoxifen therapy in recurrent breast cancer. JNCI, *87*: 751-6, 1995.

39. Miki, Y., Swenson, J., Shattuck-Eidens, D., Futreal, P.A., et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science, *266*: 66-71.

40. Wooster, R., Neuhausen, S.L., Mangion, J., et al. (1994) Localization of a breast cancer susceptibility gene, BBCA2, to chromsome 13q12-13. Science, *265*: 2088-2090.

41. Futreal, P.A., Liu, Q., Shattuck-Eidens, D. (1994) BBCA1 mutations in primary breast and ovarian carcinomas. Science, *266*: 120-122.

42. Masters, C.L., Simms, G., Weinman, N.A., Multhaup, G., McDonald, B.L. and Beyreuther, K. Amyloid plaque core protein in Alzheimer disease and Down syndrome. Proc. Natl. Acad. Sci. USA *82*: 4245-4249, 1985.

43. Mattson, M.P., Rydel, R.E., Lieberburg, I., Smith-Swintosky, V.L. Altered calcium signaling and neuronal injury: stroke and Alzheimer's disease as example. Ann. N. Y. Acad. Sci., *679*: 1-*21*, 1993.

44.. Barger, S.W., Fiscus, R.P., Ruth, P., Hofmann, F. and Mattson, M.P. Role of cyclic GMP in the regulation of neuronal calcium and survival by secreted forms of beta-amyloid precursor. J. Neurochem. *64*: 2087-96, 1995.

45. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. *162*, 156-159, 1987.

46. Liu, J., Spence, M.J., Wallace, P.M., Forcier, K., Hellstrom, I and Vestal, R.T. Oncostatin M-specific receptor mediates inhibition of breast cancer cell growth, antagonism of growth factors, and down regulation of c-myc protooncogene, submitted to Cancer Res.

47. Ozzello, L. Intraepithelial carcinoma of the breast. In Hollman, K.H. and Verley, J.M. (eds): New Frontiers in Mammary Pathology. Plenum Press, New York, 1983.

48. Barsky, S.H., Siegal, G.P., Jannotta, F., and Liotta, L.A. Loss of basement membrane components by invasive tumors but not by their benign counterparts. Lab Invest. *49*: 140-147, 1983.

49. Brown, P.W., Silverman, J., Owens, E., Tabor, D.C., Terz, J.J., and Lawrence, W.Jr. Intraductal"non-filtrating" carcinoma of the breast. Arch. Surg. *111*: 1063-1067, 1976.

50. Lagios, M. Duct carcinoma in situ: pathology and treatment. Surg. Clin. North Am. *70*: 853-871, 1990.

Table 1. Partial list of differential expressed genes in normal versus cancerous breasts identified by differential cDNA sequencing

Genes more abundant in breast cancer

Genes	EST	
	Cancer	Normal
Breast basic conserved gene	33	9
Cathepsin D	5	1
67kDa laminin Receptor	4	0
Elongation factor 1	13	5

Genes More Abundant in Normal Breast

Genes	EST	
	Cancer	Normal
Matrix Gla protein	0	8
23 kDa Highly basic Protein	3	11

Genes as Breast-Specific and Differentially Expressed

Genes	EST		
	NB[1]	BC[2]	All Tissues
BCSG1	1	6	8
BCSG2	0	7	7
BCSG3	0	5	5
BCSG4	4	0	4
BCSG5	0	4	4

[1] normal breast; [2] breast cancer

Table 1. Complementary DNA libraries were established from a stage III breast carcinoma and patient-matched normal breast. A total of 5,934 ESTs were randomly picked and sequence analyzed. More than 2,200 distinctive EST groups were analyzed for quantitative comparison of EST hits in the pair of cDNA libraries from breast cancer versus normal breast as described in "Materials and Methods". The same EST groups were also analyzed by examining the tissue-specific expression against the total of 500,000 ESTs from a variety of different cDNA libraries. Only a unique EST group with more than 3 breast-specific EST hits was listed and the rest of the several dozens EST groups with fewer than 4 breast-specific EST hits were omitted in this list.

Fig. 1. Messenger RNAs from normal and diseased tissues were extracted and used for making the cDNA libraries. These libraries are searched by EST method involving automated DNA sequence analysis of randomly selected cDNA clones. The ESTs with overlapping sequences were grouped into unique EST groups. Each unique EST group, which does not overlap to each other in sequence, was analyzed for its relative expression by examining the number of expressed individual EST in the libraries of normal vs diseased tissues. Three EST groups are listed. Blue EST group represents gene that is equally expressed in both libraries. Green EST group represents gene that is more expressed in normal library compared to diseased library. Red EST group represent gene that is more expressed in diseased library compared to normal library.

Fig. 2. Comparison of the predicted amino acid sequence with the sequence of non-Aß component of AD amyloid protein using SwissProt. After optimal alignment using the clustal method of the MegAlign Program from the DNASTAR software package, the putative protein shows 54% sequence identity with the non-Aß fragment of human AD amyloid protein.

Fig. 3. The expression of BCSG1 gene in a variety of normal human adult tissues. Twenty micrograms of total RNA from each of the above tissues was analyzed in Northern bolt using a random primer probe. A strong hybridizing band of about 1 kilobase was recognized in the lane corresponding to RNA from adult brain. A weak 1 kb transcript was also detected in testis, heart, spleen, colon, and ovary.

Fig. 4. Northern blot analysis of BCSG1 expression in human breast. Total RNAs were prepared from breast tissues and breast cancer cells and then subjected to Northern blotting analysis with ^{32}P-labeled full-length BCSG1 cDNA probe (A). The integrity and the loading control of the RNAs were

ascertained by direct visualization of the 18 S rRNA in stained gel (B). Each lane contained 30 ug of total RNA. 1: normal breast reduction mammoplasty specimens; 2: infiltrating breast carcinoma; 3: breast cancer cell CAMA-1.

Fig. 5. Northern blot analysis of BCSG1 expression in human breast cancer cell lines. Total RNA was isolated and analyzed (15 ug/lane) by Northern blot. After hybridization and washing, the filter was exposed to X-ray film for 48 hours. Lane 1: H3396 (derived from pleural effusion). Lane 2: MCF7 (derived from pleural effusion). Lane 3: SKBR-3 (derived from pleural effusion). Lane 4: MDA-MB-231 (derived from pleural effusion). Lane 5: H3914 (derived from infiltrating ductal carcinoma). Lane 6: H3922 (derived from infiltrating ductal carcinoma). Lane 7: ZR-75-1 (derived from infiltrating ductal carcinoma). Lane 8: T47D (derived from infiltrating ductal carcinoma). Cell lines of T47D, ZR-75-1, SKBR-3, MCF-7 and MDA-MB-231 are from ATCC; all other lines are initially isolated at Bristol-Myers Squibb Pharmaceutical Research Institute (46).

Fig. 6. *In situ* hybridization analysis of BCSG1 expression in human breast. Cells labelled with brown color indicate BCSG1 gene expression. All sections were counterstained lightly with hematoxylin for viewing negatively stained cells. **(A)** A highly infiltrating breast carcinoma showed a very strong BCSG1 expression in virtually every malignant cell. **(B)** High magnification view of breast cancer cell invasion to normal lobule; solid arrow indicates negatively-stained residual normal lobular epithelial cells and open arrow indicates positively-stained invasive cancer cells. **(C)** Comedo type DCIS showing BCSG1 staining. **(D)** Negative staining of BCSG1 in a non-Comedo type DCIS. **(E)**. Negative staining of normal ductal epithelial cells. **(F)** Negative staining of epithelial cells in a benign hyperplasia. **(G)** Negative staining of a benign fibroadenoma.

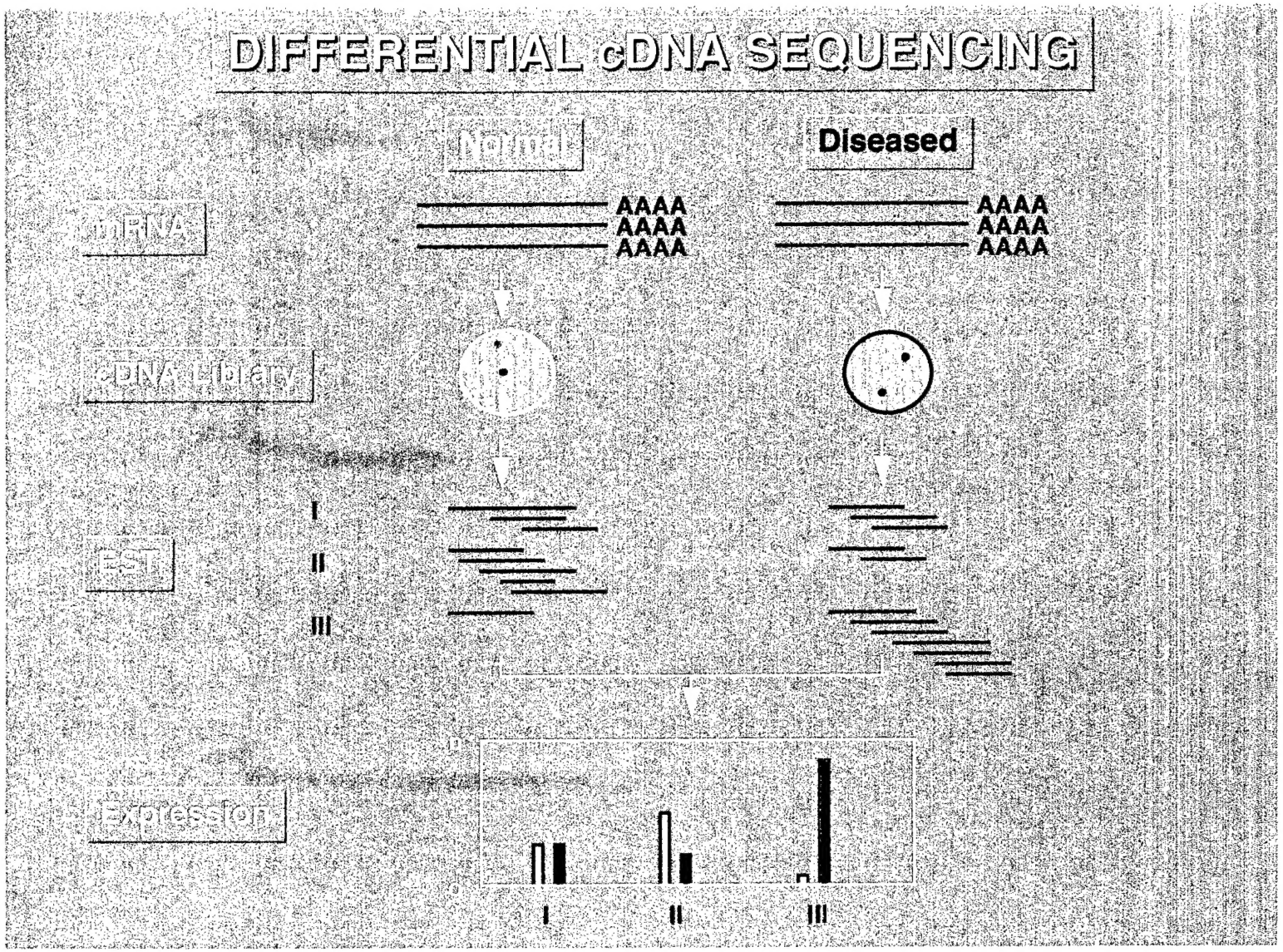

DIFFERENTIAL cDNA SEQUENCING
Normal
Diseased
mRNA
AAAA
AAAA
AAAA
AAAA
AAAA
AAAA
cDNA Library
EST
I
II
III
Expression
I
II
III

Figure 1

Fig. 2

1	M D V F K K G F S I A K K G V V G A V E	BCSG1
1	M D V F M K G L S K A K E G V V A A A E	*Human AD Amyloid
21	K T K Q G V T E A A E K T K E G V M Y V	BCSG1
21	K T K Q G V A E A A G K T K E G V L Y V	*Human AD Amyloid
41	G A K T K E N V V Q S V T S V A E K T K	BCSG1
41	G S K T K E G V V H G V A T V A E K T K	*Human AD Amyloid
61	E Q A N A V S K A V V S S V N T V A T K	BCSG1
61	E Q V T N V G G A V V T G V T A V A Q K	*Human AD Amyloid
81	T V E E A E N I A V T S G V V R K E D L	BCSG1
81	T V E G A G S I A A A T G F V K K D Q L	*Human AD Amyloid
101	R P S A P Q Q E G E A S K E K E E V A E	BCSG1
101	G K N E E G A P Q E G I L E D M P V D P	*Human AD Amyloid
121	E A Q S G G D	BCSG1
121	D N E A Y E M P S E E G Y Q D Y E P E A	*Human AD Amyloid

*Non-Aß component of Alzheimer's disease (AD) Amyloid

Figure 3

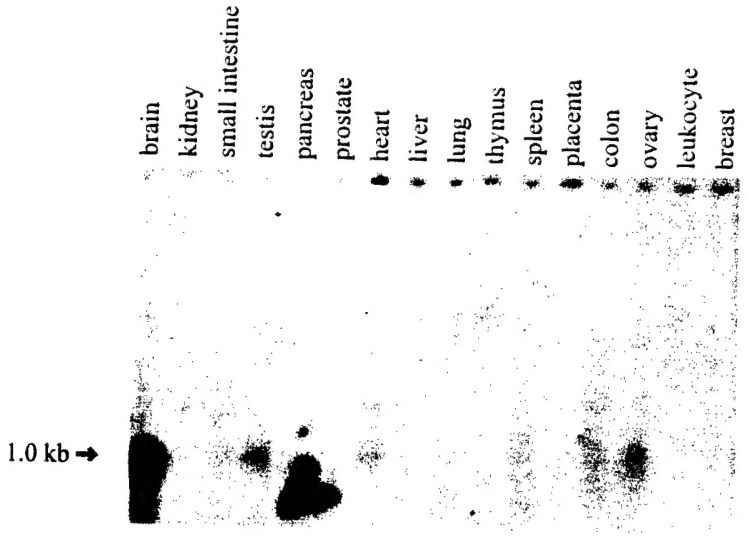


Figure 4

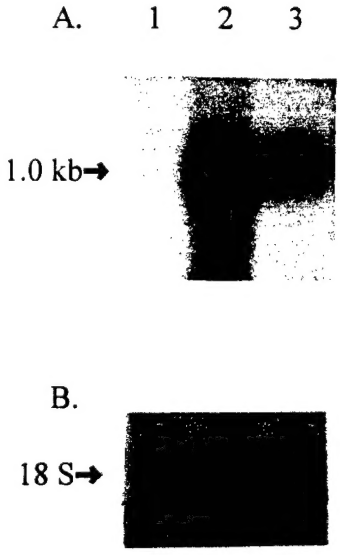


Figure 5

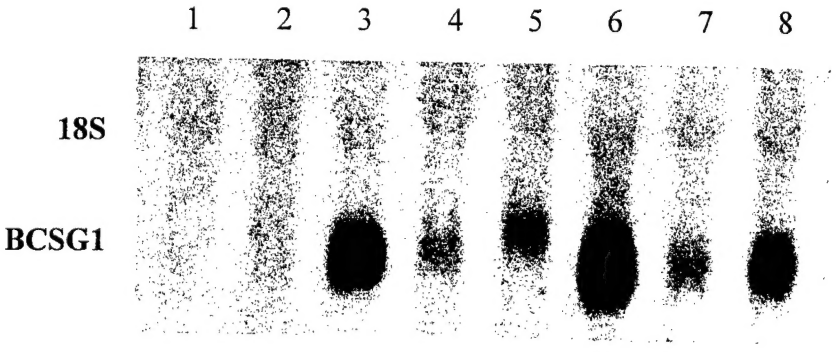


Figure 6